

In defence of the Department of Energy

Teething troubles on the US neutron spallation source project at Oak Ridge do not justify congressional threats to freeze funding, and should be viewed against a fundamentally problematic background.

The latest major civilian science project to be undertaken by the US Department of Energy is already under threat in the Congress, even though its construction has hardly begun (see *Nature* 398, 739; 1999).

There are solid grounds for concern about how the planned Spallation Neutron Source project at the Oak Ridge National Laboratory in Tennessee has been managed during its initial phase. James Sensenbrenner (Republican, Wisconsin), the chair of the Science Committee in the House of Representatives, has done his job by drawing public attention to these concerns. But in going further, and attempting to block funds for the project even as the Department of Energy takes the necessary steps to fix it, Sensenbrenner shows the lack of a sense of proportion that tends to characterize congressional oversight of the department.

Whenever trouble arises in some corner of the department's sprawling portfolio of responsibilities, its leadership and staff get the blame, both in Congress and in the country. The present Congress harbours a visceral dislike for a department that was established by President Jimmy Carter in response to the energy crises of the 1970s, and which many Republicans would like to see abolished. Alas, neither the admonishment nor abolition of the Department of Energy is likely to solve many of the problems on its patch.

In the case of the construction of large scientific facilities, for example, the root of past and present difficulties is essentially political rather than administrative. The issues that dogged the Superconducting Super Collider to death and now threaten the Spallation Neutron Source go way beyond the remit of the responsible officials at the Department of Energy. They go to the heart of the American political system — which requires, in effect, that as many states as possible get a slice of the action.

Grand design

Many years ago, under a Republican administration, the Department of Energy developed proposals for a series of large scientific facilities which, it was thought, would give US scientists the tools they would need while also securing the future of the various laboratories that would host the facilities. Under this grand design, Argonne laboratory in Illinois got the Advanced Photon Source, for example, and Brookhaven in New York got the Relativistic Heavy Ion Collider.

Under this informal carve-up — which allowed all of the host states to unite in support of all of the projects — Oak Ridge was supposed to get the reactor-based Advanced Neutron Source. But at an estimated \$3 billion, this was the most expensive facility in the series (Oak Ridge was the largest of the non-weapons laboratories) and suffered the grievous defect of being the last in the queue to get built. It was cancelled by the Clinton administration in 1995. Only then did the energy department start to develop plans for a less ambitious spallation neutron source, as a consolation prize for both the Oak Ridge laboratory and the increasingly frustrated neutron-science community.

No one ever said that Oak Ridge was ready to build a spallation source: the necessary technical expertise, in accelerator construction and neutron scattering, lay elsewhere. Also, the project needed political

support outside the state of Tennessee. So the department — encouraged by the Congress — said that five laboratories should collaborate on the construction of the Oak Ridge facility. That arrangement has proved difficult to orchestrate. The laboratories are tough to control because of their direct links to the Congress. To take the most obvious example, the energy department gets its money from a Senate appropriations subcommittee that is chaired by Senator Pete Domenici (Republican, New Mexico), who is sometimes known as the “senator for Los Alamos”, in honour of his services to the nuclear weapons laboratory in his home state. Congress likes to run things that way — almost as much as it likes to berate the energy department for failing to control its laboratories.

Problems of the territory

In the drumbeat of criticism that accompanies the Department of Energy, the difficulties of building scientific facilities comprise only a small part. This year, the strongest complaints concern alleged spying by China at the nuclear weapons laboratories. The allegations remain ill defined but, assuredly, the department will get the blame.

While it is true that the department is responsible for laboratory security, the fundamental debate between weapons scientists and intelligence people about what constitutes adequate laboratory security, which has accompanied the latest allegations, dates back to the foundation of the Los Alamos laboratory, predating the Department of Energy by a good 30 years. The laboratories have to be kept open to outside scientific influence and closed to espionage, while at the same time performing various new collaborative functions with the former enemy in Russia. Not an easy task.

But maintaining that balance at the weapons laboratories is positively a breeze next to the Department of Energy's other nuclear responsibilities, which include its mandate to receive and dispose of civilian and military nuclear waste, and to clean up the former nuclear weapons production complex. The latter task is being undertaken by energy department officials who are under continual siege not just from irate congressmen and state governments, but also from lawyers making use of complex environmental statutes to attack any selected course of action from every conceivable angle. The constraints that dictate the progress, or lack of it, in this, the world's largest environmental restoration project, were set not by the department, but by the Congress.

There is, of course, ample room for improvement in the Department of Energy's performance and in its leadership. But the toughest problems the department faces aren't attributable to maladministration. Rather, they come with the territory.

The Spallation Neutron Source project at Oak Ridge has got off to a shaky start. If construction money is halted by the Congress at this early stage, the project will quickly enter a spiral of delays and cost-overruns that will probably destroy it. There is no need at all for this to happen. The project has a much-respected new director, David Moncton, who should be given a chance to install the right team and integrate the activities of the five laboratories. With his management and the support of the Congress, the project can be completed on time and on budget. □



Clinton pledges to smooth path of research at universities

[WASHINGTON] President Bill Clinton pledged last week to “renew the alliance between America and its universities”. He endorsed an action plan produced by his administration which, he says, will attract more women and minorities into science, “cut down the red tape” confronting researchers and clarify the status of graduate students.

“America’s scientists should spend more time on research, not filling out forms in triplicate,” said Clinton. He was speaking at a ceremony where he presented national medals of science and technology to scientists, engineers and corporations.

He has accepted a plan proposed by his National Science and Technology Council (NSTC), a committee of all the main science administrators in the federal government. This follows a three-year study of trouble spots between universities and government.

“We must move past today’s patchwork of rules and regulations and develop a new vision for the university–federal government partnership,” he said.

The NSTC promises that government science agencies will agree on changes to “reduce differences” in the way they administer research grants, and implement them within 12 months. This should make it easier for researchers to apply for funding to different agencies without repeating themselves.

The plan also promises to streamline requirements for grant recipients to certify that they meet certain legal standards, such as providing a workplace free of recreational drugs. Changes could enable certification to be obtained on an institution-wide basis, says NSTC, greatly reducing paperwork.

The study acknowledges the extensive ad hoc demands for cost-sharing with the universities that some federal agencies are placing on grant recipients — to the alarm of universities. It promises to get agencies to state clearly what their policies are on cost-sharing, again within 12 months.

And the NSTC promises to review policies under which some parts of the government, notably the Internal Revenue Service, refuse to accept the status of graduate students as both employees and students.

However, the NSTC study calls for more talk — particularly with the universities — rather than for specific action to be taken now to tackle any of these problems. Similarly, a memo issued by Clinton in response to the study calls for recommendations to be developed within 12 months, rather than ordering immediate changes in policy.

Still, the universities hailed the proposals and Clinton’s endorsement of them. “We



Clinton awards cancer researcher Janet D. Rowley a national science medal at last week’s ceremony.

welcome the fact that the president of the United States has placed this on his agenda,” says Nils Hasselmo, president of the American Association of Universities, which represents 50 leading research universities.

Especially helpful, says Hasselmo, are the promises to clarify the status of graduate students and to spell out policy on cost-sharing.

“Cost-sharing too often becomes a matter of negotiation between the programme director and the principal investigator, and that has raised grave concerns,” he says. “We want to make cost-sharing fair and effective, as well as transparent.”

The administration plans half a dozen public hearings around the country, starting

in Washington at the end of this month, to discuss its proposals with university administrators and researchers.

But some observers, while welcoming the ambitious scope of the review, worry that it is taking place too late in Clinton’s second term to lead anywhere. “It has some potential if the administration stays focused on it,” says one official familiar with university–government relations. “If Al Gore is elected in 2000, maybe it will carry over.”

During the award ceremony last week, Clinton offended Republican members of the House Science Committee by calling Rush Holt (Democrat, New Jersey), a physicist who was elected to the Congress last November, “the only bona-fide scientist in the US Congress”. “It is about time we had one and I’m glad he’s here,” added the usually well-briefed president.

His remark wounded at least two Congressmen who had been working on science issues for years before Holt’s election. Vernon Ehlers (Republican, Michigan), who has just written a National Science Policy Study for the Congress, is a former physics professor; Roscoe Bartlett (Republican, Maryland) is a physiologist and inventor with 100 published papers and 20 patents to his name.

“These two fine leaders might not have created the Internet,” said James Sensenbrenner (Republican, Wisconsin), chair of the Science Committee, in a reference to Vice-President Al Gore’s recent paternity claim on the medium. “But in my book they are bona-fide scientists.” **Colin Macilwain**

UK failing to reap neuroscience rewards

[LONDON] British companies are failing to reap substantial commercial benefits from neuroscience, even though British research is the second most cited in neuroscience patents in the United States, according to a 12-country audit of such research just published by the Wellcome Trust.

The audit reveals that, although 13 per cent of inventors named on US neuroscience patents are British, only 3.5 per cent of these patents are owned by British individuals or companies. Ownership is dominated by US industrial corporations.

The audit says this ‘exploitation gap’ in neuroscience — a generic problem in UK academic research — needs to be addressed. One helpful measure, it says, is the £60 million (US\$97 million) ‘University Challenge’ fund set up by the government, the trust and the Gatsby Foundation to help

universities commercialize their research.

The audit analysed 322 neuroscience research papers cited in 371 US patents. Three quarters of papers were basic research as opposed to clinical studies. Britain’s 110 named inventors came second only to the United States, which dominated with 561. France had 50, Japan 29, Italy 20; Germany came third from bottom with two inventors.

Researchers from the United States dominate eight subfields in neuroscience by publishing around 50 per cent of papers in basic research and 60 per cent of clinical studies. Britain, however, is the only country whose share of research is steadily increasing in most neuroscience subfields.

Neuroscience research represents currently 16 per cent of the 227,000 papers in biomedicine published annually around the world. **Ehsan Masood**

France seeks scientists' views on reform

[PARIS] Prime minister Lionel Jospin launched a national consultation on the organization of French research last week. He has commissioned the exercise after earlier reform plans by science minister Claude Allègre stalled in the face of opposition from the research community.

A parliamentary mission will culminate in a national debate in Paris in June, and is expected to present Jospin with "concrete proposals" the following month. The mission is led by two socialist members of the national assembly, Pierre Cohen (Haute-Garonne) and Jean-Yves Le Deaut (Meurthe-et-Moselle), a former head of the Parliamentary Office of Scientific and Technological Choices.

Tension over Allègre's plans is still running high (see *Nature* 397, 463; 1999). The main unions representing scientists suspect that the ministry may simply pay lip service to the outcome of the consultation — Allègre had previously rejected demands by researchers for a national public debate.

Last week the unions called on researchers to support a national day of protest on 11 May, to secure a commitment

from the government that no reforms would be undertaken before the end of the consultation. In particular, they are demanding the postponement of an interministerial meeting on research due to be held on 18 May. They argue that the decision not to defer this meeting suggests that the ministry is likely to ignore the consultation and push ahead with its own reform plans.

But Vincent Courtillot, director-general of research at the ministry, dismisses such concerns. He claims that the interministerial meeting will deal only with strategic research themes and industrial research.

"The CIRST [conseil interministériel sur la recherche scientifique et technologique] will not be centred on the items included in the parliamentary mission, though it is likely to mention that actions on helping young researchers and promoting scientific mobility will be taken after the Cohen-Le Deaut report has been submitted," says Courtillot.

Similarly, Le Deaut says that he would not have accepted the mission had he felt it was a tactic by the government to defuse opposition. He says he has assurances that Jospin

will take the outcome seriously and that it will not "gather dust in a drawer".

The mission is being assisted by a steering committee of more than 20 researchers and industrialists, including former science minister Hubert Curien and Philippe Desmarescaux, director-general of Rhône-Poulenc.

In a break with tradition, the committee also includes several young scientists. Finding ways to give young scientists greater independence is one of the five stated goals of the mission, and the theme of a debate in Paris on 9 June.

The others goals are to propose ways of increasing the mobility of researchers; to improve links between research agencies and universities; and to analyse procedures for recruitment and evaluation.

In hearings being held over the next few weeks, the mission will consult about 100 leading research figures, while wider input will be sought using an Internet forum (see <http://www.mission-cohen-ledeaut.org>). A series of public meetings in various French cities is also to be organized.

Declan Butler

Germany says it will come on board Earth observation mission

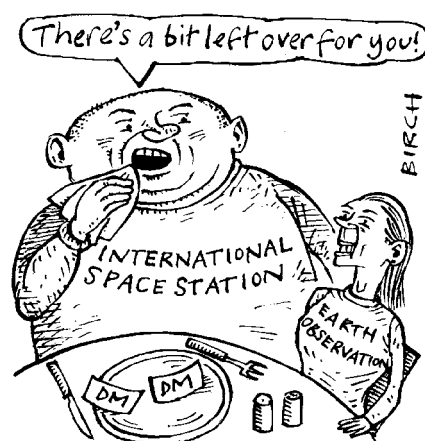
[MUNICH] German research minister Edelgard Bulmahn announced last week that the country will participate "substantially" in the European Space Agency's (ESA's) 'optional' Earth observation and launcher development programmes.

Her announcement has eased fears that Germany's contribution, which could make or break the programmes, might not emerge. But Bulmahn has not yet revealed how much Germany will contribute — that sum will be negotiated at next week's meeting of European space ministers.

The ministerial meeting has been delayed for nearly a year because of Germany's financial problems. It is being held to discuss the levels of resources and broad content of ESA's programmes over the next five years.

Bulmahn has long complained that heavy commitments to the International Space Station, which the Social Democrat-Green coalition government inherited from its predecessor, left no spare money for joining in new, unmanned space activities. These include Earth observation, one of the government's highest scientific priorities.

But, after a detailed review of potential savings in each part of the German space budget, Bulmahn says she has found room for manoeuvre. In addition, she says, the overall German space budget will be raised over the next few years, and she expects to win a DM10 million (US\$5.4 million)



increase to the 1999 budget of DM1.6 billion over the next few months.

Money for space activities is tight in other ESA member states as well. So prospects are bleak for an increase in resources for the 'mandatory' science programme, which ESA's director-general Antonio Rodotà calls the "backbone" of the agency.

Resources have been capped for the past four years, squeezing budgets for approved missions almost to breaking point (see *Nature* 395, 732; 1998). Rodotà will ask ministers to approve an internal reallocation of funds to maintain 1998 purchasing power until 2003, to ensure that all planned missions are launched on schedule.

Following Germany's announcement,

ESA officials expect to be able to raise 759 million euros (US\$804 million) at the ministerial meeting for the optional Earth observation programme, which has been redesigned on the 'faster, cheaper, better' principle (see *Nature* 392, 425; 1998).

The programme has a strong scientific element, with its Living Planet component aiming to achieve a launch every year. Small scientific missions will be interspersed with medium-sized 'core' scientific missions.

Selections for the first of both types of mission are currently being made by the scientific community. Front runners include a small ice altimetry mission called Cryosat, which will measure ice topography and ice motion, and a 'core' gravity mission called GOCE (Gravity Field and Steady State Ocean Circulation Explorer), which will measure irregularities in the Earth's mass distribution through gravity-field variations, with the aim of better understanding ocean circulation.

Space ministers will also discuss telecommunications and launcher plans next week. The latter include an extension of Ariane V and the development of smaller launchers to maintain international competitiveness. Ministers will be asked to finally approve plans to develop a second-generation Global Navigation Satellite System in partnership with the European Commission, at a cost to both partners of 500 million euros.

Alison Abbott

Berkeley dispute festers over biotech deal

[SAN DIEGO] A \$25 million plant research contract with the life-sciences company Novartis is continuing to stir controversy at the University of California at Berkeley (UCB).

Some faculty and graduate students at the university's College of Natural Resources are deeply concerned about the arrangement's propriety and its potential impact on academic freedom. They are planning to ask the state legislature to address the issue.

The contract took effect last November (see *Nature* 396, 5; 1998). Its supporters, and Novartis spokespeople, say the agreement is a great advantage for the college, faculty and students. College officials say they are unaware of any inappropriate activity associated with the contract.

But critics say academic research may already be inappropriately affected. They point out that at least one graduate student has left a doctoral programme in part because of the deal. They claim that both students and faculty members feel intimidated about voicing their concerns.

"Before the Novartis deal, we had a college with diversity of opinion; now we have a divided college," said Ignacio Chapela, an assistant professor of microbial ecology, who is the elected chairman of the executive committee of the college's faculty.

The college's mainstay is research on plants and agricultural and ecological policy, in one of the world's most intensely farmed states. It is divided into four departments, including the department of plant and microbial biology that under the terms of the contract will receive \$5 million a year for five years from Novartis.

Two-thirds of the money will go into research, with one-third for indirect costs at the university and department. Much of the Novartis-funded research is expected to involve genetically modified plants and associated fields. Faculty and students are to get access to Novartis's genomic databases of information on plants.

In return, Novartis will have first right to negotiate for licences on a percentage — about 30 per cent in the first year of the contract — of patentable discoveries coming from research carried out by individuals in the participating department. Twenty-nine of 32 faculty in the department have signed up to participate in the agreement. The university will retain ownership of the research results and patents.

Gordon Rausser, the dean of the college and principal architect of the Novartis deal, calls the critics' concern "silly". Regarding any move to have the state legislature examine the deal, Rausser says: "I would be delighted with any scrutiny; it will prove we enhanced and leveraged public resources."

In recent months displeasure over the



Up for sale? Some at UC Berkeley fear business deals may compromise academic freedom.

Novartis deal has erupted in a number of Berkeley campus quarters, including an article in the February issue of the UCB alumni magazine, *California Monthly*, by Robert Berring, the Walter Perry Johnson Professor at UCB's School of Law.

"We must ask at what point does the university bargain away so much of itself that it ceases to be the university and becomes a partner of the private sector?" wrote Berring in the article. "If the private sector begins to play a role in setting research priorities, if it gets built into the budget, something very fundamental has changed."

The current issue of the magazine

includes an article by Rausser, supporting the Novartis deal, that is seen as a sharp response to Berring.

"In the final analysis, (UC) Berkeley is a public asset of immense value," writes Rausser. "So long as our culture is maintained, this value will be enhanced, not diminished, when we work creatively in collaboration with other institutions, including private companies."

There is some evidence that the nascent Novartis deal is already affecting the university culture. Susan West, a doctoral candidate at the college, claims that the Novartis deal was "the clincher" in prompting her to drop out of the programme after completing her first term in December.

"I was uncomfortable with it," says West, who was pursuing research into arctic terns after completing a Fulbright Fellowship in Iceland. "It impugned the entire college. Do they really believe that [\$25 million] will not swing the balance?"

There are unconfirmed reports that, during a recent oral exam on a doctoral student's thesis, the student was encouraged to pursue a line of research with Novartis. Several college faculty have said that this sort of intrusion into the academic process, if it did take place as described, was clearly inappropriate. Rausser says that he would be concerned about such an interaction, but that he is unaware of any such report.

Rex Dalton

Embattled neuroscientist wins US support

[MUNICH] Forty-six top US neuroscientists have written an open letter to politicians and university professors in the German state of Bremen protesting at the treatment of a leading neuroscientist who has received a string of threats from animal-rights activists.

The signatories to the letter include the Nobel laureates Francis Crick and David Hubel. They say that they endorse the basic research into cognition that Andreas Kreiter is carrying out using macaque monkeys, and stress that the work could have long-term importance for treating psychiatric and neurological disorders.

The letter also attacks the lack of local support for Kreiter, who has been continually criticized both by animal rights activists and by some of his colleagues since his arrival in Bremen in 1997 (see *Nature* 395, 505; 1998).

"It is perhaps understandable that some individuals ... can be influenced by the narrow ideology and polemics of the animal rights organizations," says the letter. "We are astonished, however, that substantial numbers of academicians, government representatives and journalists appear

sympathetic to the efforts of these organizations to prevent Professor Kreiter and others from performing their research."

Last year, more than 100 professors at the University of Bremen — including a number of biologists — signed a letter calling for Kreiter's research to be terminated. The letter was sent to all members of the university members and to the government of Bremen.

But Hans Flohr, director of the University of Bremen's brain research institute, says the letter is a "great help". He adds: "In a small state like Bremen, politics is very close to the people, and politicians feel very much the influence of pressure groups."

The powerful international voice in support of Kreiter's research is a strong counterbalance to this pressure, he says.

Wolfgang Appel, president of the German Animal Protection Association, says he understands that scientists working in the same field would want to support each other, but says he could gather "hundreds of other scientists who could give equally well-argued cases against the value of Kreiter's primate work".

Alison Abbott

Global R&D spread clouds local analyses

[PARIS] The 'globalization' of investment in industrial research and development (R&D) is complicating analysis of the link between a nation's public science base and its economic competitiveness. The notion of individual companies acting as 'national champions' has been rendered obsolete in the process.

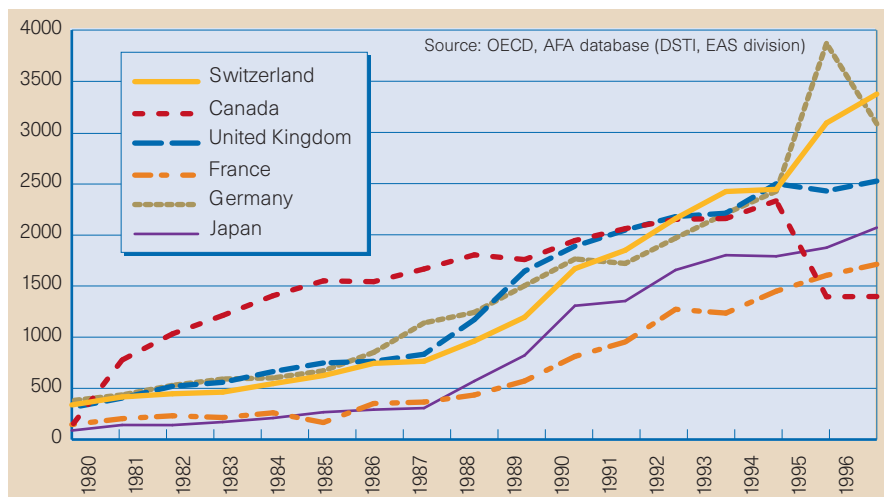
These conclusions emerge from a report from the Organization for Economic Cooperation and Development (OECD). The report reveals that, between 1977 and 1994, foreign investment in the United States increased from US\$933 million to \$15.6 billion (see chart). Most of this increase was accounted for by European companies, whose investment jumped from \$790 million to \$10.3 billion.

The same trend is apparent worldwide, says the report, which is the first major survey of the internationalization of industrial R&D. "It constitutes a very important document," says Nicholas Vonortas, associate professor of economics and international affairs at George Washington University's Center for International Science and Technology Policy. Vonortas says that the report brings together the first real data on the issue. "I think it will spur a new wave of much-needed empirical research on the subject."

One result of the growth in foreign R&D investment is that "it is becoming more difficult to assess what is 'national' industrial R&D, in our traditional perception of what is national", says Thomas Hatzichronoglou, the report's principal author. Hatzichronoglou is a senior official at the economic analysis and statistics division of the OECD's Science, Technology and Industry Directorate.

For example, whereas national league tables were relatively easy to establish when industrial research was dominated by 'national champions', the domination of multinational companies means that "it is increasingly difficult to establish who benefits from foreign investment in R&D", he says.

The report challenges the wide belief among economists that R&D is one of the few industrial operations to have escaped globalization, says Hatzichronoglou. Despite a clear economic case for decentralizing production



On the rise: increase in R&D expenditure (US\$ million) in the United States by foreign companies.

facilities close to markets and sources of cheap labour, there is a general perception that companies prefer to keep strategic R&D operations at home, he says.

"But if you look at the data this is not true," says Hatzichronoglou. The globalization of R&D is "much more pronounced than people believed, and the trend is accelerating".

Keith Pavitt, of the Science and Technology Policy Research at Britain's University of Sussex, cautions that investment by foreign subsidiaries — at around 12 per cent of total industrial R&D spending — is small compared with production and marketing. "It remains one of the least globalized of all business activities."

Pavitt says that the main question raised by the report's findings is the nature of the relation between a country's national science base and its welfare. "The past assumption of a flow from national laboratories to national companies is increasingly outdated," he says.

Competitiveness "does not necessarily flow through national companies", says Pavitt. He believes that more sophisticated thinking is needed on national industrial policy. "The links between the public science base and production are much more complicated [than often assumed]," he says.

One reason for globalization is that multinationals increasingly need to tailor products

— particularly drugs — to local market and regulatory conditions. But much of foreign industrial R&D investment is aimed at supporting worldwide markets, and companies seem to be seeking out research talent in universities and institutes. At the same time the Internet — and intranets — have lifted many of the logistical obstacles to managing multi-centre research programmes on a global scale.

According to the OECD report (*Internationalisation of Industrial R&D. Patterns and Trends*, 1998), in addition to alliances between companies and universities, the increase in foreign investment is strongly driven by mergers and acquisitions of foreign companies possessing significant research activities.

But Hatzichronoglou says the report shows that concern among countries over the transfer of R&D activities abroad by domestic companies is exaggerated. Although there are insufficient data for an unequivocal answer, the indications are that it is the parent company that profits most from decentralization, he says.

He points, for example, to the benefits of repatriated profits and skills and the improved knowledge of local markets. "Relocation is often badly perceived but ultimately it is the company that benefits."

Host countries clearly gain from attracting foreign research investment, in that this encourages high-technology jobs — 105,100 US jobs in 1994 were a result of foreign R&D investment — training, and technology transfer. But Hatzichronoglou believes that while host countries benefit, "the parent group is the main beneficiary", particularly in terms of patents.

Pavitt points out that there is no substitute for a strong domestic science base. But in the absence of this, a foreign presence is better than the alternatives — persisting with an inefficient domestic industry, or licensing in technology.

Declan Butler

Africa moves to strengthen biotechnology

[LONDON] Governments in Africa have set up an agency to help develop biotechnology across the continent. The African Agency of Biotechnology will hold its first regional workshop in September in Algiers, where the new agency is based.

According to its director-general, Samuel Nzietchueng, the agency is designed to encourage African states to develop

biotechnology, both as a tool to meet future food needs and as a means to create wealth from commercializing research.

The agency's aims include coordinating research, and encouraging countries with a stronger base in biotechnology to help weaker neighbours. Other objectives include harmonizing regulations on bioethics and intellectual property rights. □

GM advisory panel is slanted, say critics

[WASHINGTON] The National Research Council (NRC) in Washington has come under fire over allegations that a panel set up to examine genetically modified pesticide-producing plants is slanted in favour of the biotechnology industry.

The panel is investigating the risks and benefits of such plants and examining their federal regulation. The NRC, an arm of the National Academy of Sciences, received 320 letters on the panel's composition in a public comment period that ended last week. Most of these — 268 — were negative, with many complaining that the 12-member group includes too many people sympathetic to the biotechnology industry.

In a letter to the NRC, the Union of Concerned Scientists (UCS) and six other groups say that the panel is "weighted with molecular and agricultural scientists", to the exclusion of specialists in food allergies, soil microbiology and genetic outcrossing to wild and weedy relatives of engineered crops. "[The panel] wasn't even close in terms of balance," says Margaret Mellon, director of agriculture and biotechnology at the UCS.

The NRC plans to respond to the letters this week by announcing the appointment to the panel of Rebecca Goldberg, a senior

scientist at the Environmental Defense Fund. It will also announce the resignation of Brian Staskawicz, a professor of plant biology at the University of California, Berkeley.

Staskawicz was a signatory to a controversial \$25 million deal between the university and Novartis, in which the life sciences company is funding research in his department (see page 5). He resigned voluntarily last month, says E. William Colglazier, NRC executive officer.

"Since we [at NRC] are only advisory... the only thing that really counts is our credibility, how objective we are," says Colglazier. "We take that very seriously [in] trying to put together a very expert committee" that is balanced and deals seriously with potential conflicts of interest, he says.

Colglazier says that the remaining members of the panel are being vetted for the extent to which their income is generated from sources that have a vested interest in the outcome of the study.

Critics have also questioned the current interests of two former officials of the Environmental Protection Agency (EPA) who are on the panel. One, Stanley Abramson, now works at a major Washington law firm representing companies seeking approval for

genetically engineered agricultural products and defending products before government agencies and the courts.

The other, Fred Betz, represents biotechnology companies applying to EPA for registration of plants genetically engineered to produce pesticides. Betz says his involvement does not entail an "irreconcilable" conflict of interest. "Quite the contrary," he says. "The fact that [my] firm is working with the industry and with EPA [puts me in] a unique position to have a perspective on what's working and what may not be working in terms of the regulatory framework."

The NRC report is being drafted at a time when the EPA is finalizing the regulations for controlling the registration of pesticide-producing genetically modified plants. A coalition of 11 scientific societies, including the American Society for Microbiology, complained vociferously in 1996 that the EPA policy was "scientifically indefensible" in targeting genetically engineered plants for regulation (see *Nature* 382, 485; 1996).

The NRC says that the 1996 report helped spur it to undertake the study. It is financing the study from its endowment, in contrast to most of its studies, which are paid for by the government.

Meredith Wadman

Australians seek international allies in battle over uranium mine

[SYDNEY] Australia's environment minister, Robert Hill, has launched an international battle to stop Kakadu National Park being named an endangered World Heritage site. The dispute centres on Jabiluka uranium mine, which is being developed — with Hill's approval — on the park's border.

Last week, the small Australian Democrat party led a Senate defeat of the Coalition government, by winning approval for a committee of inquiry to investigate the government's process for approving the Jabiluka project. Democrat Senator Lyn Allison, chair of the committee, says the environmental impact assessment Hill used lacked independent peer review by scientists.

According to leaked documents, Hill is exerting diplomatic pressure on UNESCO member countries to vote against the 'endangered' recommendation when the full World Heritage Committee meets in July.

Hill's campaign is based on challenges to assessments by two UNESCO advisory bodies: the World Conservation Union (on environmental effects) and the International Commission on Monuments and Sites (on cultural factors). The credibility of the UN agency itself is also being questioned (see *Nature* 396, 606; 1998 & 397, 287; 1999).

He is relying on a rebuttal of the findings of a group of scientists from Australian



Mine of disinformation? Opponents question the accuracy of each other's reports on Jabiluka.

National University (ANU), led by hydrologist Bob Wasson. The ANU group had found that there was a significant risk, during unusually severe rainfall, of leakage into park wetlands of radioactivity from ore stockpiles, and from dams planned to contain tailings from the milling of ore.

The rebuttal Hill is citing criticizes the ANU conclusion and claims that "natural values" are not threatened. The unsigned report was prepared "with assistance from experts" at the Commonwealth Scientific and Industrial Research Organization, the Universities of Melbourne and New South Wales and the Bureau of Meteorology.

But the main thrust of Hill's criticisms has been directed at local Aborigines, the Mirrar people. They oppose drilling of the inclined shaft of the mine towards the rich ore body, claiming that this threatens their culture, as it approaches a sacred site. Hill has dismissed them as traditional owners of "a tiny fraction" the park's 20,000 square kilometres. He says that their wishes "would detrimentally affect the majority".

Like Hill, Aboriginal and environmental activists are campaigning internationally. Two weeks ago the A\$100,000 (US\$65,000) Goldman Prize for environmental advocacy was awarded to senior traditional owner Yvonne Margarula and to Jacqui Katona, executive officer of the Mirrar organization, the Gundjehmi Aboriginal Corporation.

Meanwhile, Hill and science minister Nick Minchin have approved the development of Australia's fourth uranium mine, at Beverley in South Australia, the first to use the process of in-situ leaching.

The government's claim that "the waste will remain isolated from the biosphere throughout time" has been contested by the Australian Conservation Foundation. The foundation says similar uranium mines in Europe and the United States "have resulted in widespread adverse groundwater impacts and contamination".

Peter Pockley

Mixed response to NIH's web journal plan

[PARIS & WASHINGTON] A proposal by Harold Varmus, director of the US National Institutes of Health (NIH), for a global web site — E-Biomed — that would centralize much of the biomedical literature and make it freely accessible has met with a mixed reaction from publishers and scientists worldwide.

Under the proposal, articles could either be submitted for publication without refereeing — as e-prints — or after peer review by third-party 'editorial boards' (see *Nature* 398, 735; 1999). The operation would be run by an independent governing body.

The goal of making the literature more readily available has been widely welcomed. But many are sceptical of the idea's desirability and feasibility. Some criticisms stem from a defence of vested interests or the status quo, while others reflect the uncertainties surrounding what is a preliminary proposal. In any case, Varmus has achieved one of the goals of the proposal: "to accelerate much-needed public discussion of electronic publication in the United States and abroad".

Axel Kahn, editor of *Médecine et Sciences*, the leading French-language biomedical journal, says the proposal challenges the "naked emperor" of scientific publishing —

that "80 to 90 per cent of what is published is of little real interest". Publish or perish "rather than intrinsic merit" has become "the principal justification for much of the output," he says.

Kahn claims that most journals are infrequently consulted, and that E-Biomed would "allow you to have access to the articles you want, without having to browse hundreds of journals". A shake-out of the journals system is long overdue, he says, adding that there is only a real need for the cream of journals, and in particular the best multidisciplinary journals.

Geert Noorman, managing director of Elsevier Science's Life Sciences Division, also says he welcomes the proposal. The exponential growth in scientific output means that "new ways of processing and organizing information are needed," he says. Noorman's comments surprise some observers, who predict that the big commercial publishers will mount a lobbying offensive to try to block public funding for E-Biomed.

One major concern is that the proposal could harm the best existing journals, without accelerating improvements that might gradually occur in any case as a result of market forces and more diffuse web efforts by

not-for-profit science publishers. Several observers say it might create an unhealthy monopoly, erode the diversity of existing journals, and reduce competition between journals for the best papers.

The Varmus proposal notes that the current journal structure has served the biomedical community well for 300 years. "So the first question I ask is, if it has served us well for 300 years, why change?" says Martin Frank, executive director of the American Physiological Society, which publishes 14 journals and 36,000 pages of articles each year.

"It [E-Biomed] is extremely cumbersome and is not going to be easily implemented," says Frank. "It is so unclear in terms of process that it's going fall under its own weight."

Frank and other non-profit publishers are irritated at what they claim has been their omission from early discussions of the proposal, even though it intimately affects them. Thirty non-profit publishers wrote to Varmus on 29 March, as word of the proposal began to spread, asking him for a meeting to discuss the plan. Varmus points out that a series of meetings is being scheduled with organizations worldwide such as the European Molecular Biology Laboratory (see box opposite), but that these will take some time to arrange and conduct. He also intends to post the proposal on his NIH web page for comment.

Some observers note that the page charges collected by some publishers provide them with a cash cow — and that in the United States the NIH is one of the largest that is milked. Under the existing system, page charges can be passed on to the biomedical agency by investigators that it supports. The potential loss of this lucrative system is alarming publishers — especially non-profit organizations — which rely heavily on it.

Proposals that E-Biomed should coordinate peer review of its contents are controversial. Noorman argues that centralization of peer review would threaten the diversity of schools of thought provided for by journals.

This concern is shared by many scientists and learned societies, who feel that a centralized structure may obscure the well-defined hierarchy of best science provided by journals, and that scientists may be more reluctant to give their time and energy free to a central structure.

Andrew Odlysko, a mathematician at the AT&T telecoms corporation and an expert on scholarly publishing, argues that it would be simpler to separate the distribution and peer-review functions of the repository, as is done at the Los Alamos physics e-print servers, where peer review is provided by journal 'overlays' to unrefereed papers.

Lynn Dobrunz, a postdoctoral neurobiologist at the Salk Institute in San Diego, asks: "Would E-biomed be in addition to the

EU warns on growth-hormone cancer risk

[WASHINGTON] A growth-enhancing hormone fed to cattle "has to be considered as a complete carcinogen", based on a substantial body of recent evidence, according to a committee of the European Commission (EC) reporting on Monday (3 May).

The EC's Scientific Committee on Veterinary Measures issued its finding at a critical time — just ahead of the 13 May deadline set by the World Trade Organization for European Union compliance with the WTO's ruling of 1997 that the EU must accept hormone-raised beef for import. In that ruling, the WTO declared illegal a 1989 EU ban on imports of hormone-raised beef, saying that the EU had not conducted a proper risk assessment. The study was meant to address that concern.

The committee report — which is interim and open for public comment — finds that 17 β -oestradiol "exerts both tumour initiating and tumour promoting effects". The report is a risk analysis of residues in beef from six growth-promoting hormones fed to cattle. It found that, for all six hormones studied, endocrine, developmental, immunological, neurobiological, immunotoxic, tenotoxic and carcinogenic effects "could be envisaged", although such risks could not be quantified.

But the study drew fire in Washington, where some argued that it was a European ploy to avoid WTO sanctions if the EU fails to accept hormone-raised beef imports.

A State Department official calls the report's timing "highly suspect". "After all these tests over all these years [showing the safety of such beef] there is suddenly a dramatic new scientific discovery... What they are desperately trying to do is find some excuse [to avoid WTO sanctions]. It's transparent," says the official.

And an official at the US Food and Drug Administration dismisses the report as void of "new information". Bert Mitchell, acting deputy director of the agency's Center for Veterinary Medicine, says: "We have very carefully reviewed the safety of these hormones over several years. Our position is that they are safe."

But the report drew praise from consumer advocates. "This is going to be a good test of the ability of Europe to withstand the pressures of an unaccountable World Trade Organization," says Jeremy Rifkin, president of the Foundation on Economic Trends in Washington. "There are serious potential health risks that should be looked at. It's not resolved that this is safe. The burden should not be on the public to prove it's safe."

M. W.

current system of journals, or instead of it? If there was a consolidated site that published online versions of all the articles that are currently published... that would be fantastic. If it's instead of, and especially if it has this non-peer-reviewed track to it, I think that is a much less good idea."

The Varmus proposal suggests that scientific societies could be one source of peer review. But the societies are worried that E-Biomed may undermine the journal revenues on which many of their other activities, such as fellowships and meetings, depend.

The head of one society says he is open to change, but would need guarantees that revenues would be preserved. Given such guarantees, societies might consider joining the initiative, he says. "E-Biomed will only fly if learned societies and their journals can be brought to the table," predicts Tony Delamothe, web editor of the *British Medical Journal*, and a supporter of E-Biomed's goals.

Another broader threat, expressed by many scientists, is that NIH might come to dominate much of the biomedical literature, leading to homogenization or to discrimination against scientists from smaller countries. "Who would select the governing body?" asks an official at one European scientific society. "Who would select the editors and decide what is allowed to be published? Who will determine costs and access rights?"

Many are also uncomfortable with the prospect of public funding for scientific publishing, an activity currently dominated by for-profit and non-profit publishers in the private sector. At the same time, however, there is growing resentment among scientists and librarians at the spiralling inflation in journal subscriptions.

Competition between scientific publishers is less than in other industries because of distortions in the market, and profit margins as high as 40 per cent are not uncommon (see *Nature* 397, 195–200; 1999).

Graham Cameron, head of services at the European Bioinformatics Institute (EBI) in Cambridge, England — an outstation of the European Molecular Biology Laboratory — points out that public domain databases such as PubMed and GenBank and the EBI are widely considered to provide high levels of cost-effective service to the community.

Many believe, however, that the wider and cheaper access promised by E-Biomed may happen anyway as a result of market forces. "Most scientific society publishers are already doing what Varmus is proposing," says Frank. "We are putting our journals on the web. We are linking our journals through PubMed to our sister journals on the web. We are developing interfaces for the submission and review of manuscripts on the web." Similarly, consortia of library and other users are increasingly negotiating electronic licences for journals for entire institutions and even countries. Scientists at such institutions can already access much of the literature online.

"My initial reaction to E-Biomed is, 'so what?'. Virtually every library has almost all major journals," says Heinz Steiner, a neuroscientist at the University of Tennessee College of Medicine in Memphis.

Market forces are also driving a flurry of deals among publishers that may enable researchers to move rapidly and seamlessly from a citation to full text across journal boundaries.

Frank asserts, for example, that the web site of HighWire Press already accounts for a large proportion of the biomedical literature. This not-for-profit outfit was set up in 1995 by Stanford University Libraries and Academic Information Resources to help universities and societies to publish on the web at low cost. "So I don't know why we need to create E-Biomed," says Frank.

Indeed, the head of one scientific society argues that resentment over the huge costs of the current journals system is confusing the many complex issues involved in scholarly publishing. "If publishers are charging too much then we should attack this problem directly, but not attack the entire system. E-Biomed is a not very selective nuclear bomb."

Noorman, while admitting that Elsevier's profit margins "are higher than the average," says that the arrival of web publishing is putting pressure on commercial publishers. "Scholarly publishing will become a proper [not distorted] market," he predicts. "Elsevier is not in the world to keep that profit margin high. We are in the world to stay in the market. If the web causes us to have to agree to lower profit margins, then so be it." **Declan Butler & Meredith Wadman**

Automation 'could crack the big problems in science'

[WASHINGTON] A new type of laboratory that would collect and analyse vast amounts of data to solve complex biological problems is being proposed by scientists from US universities and government laboratories.

The approach, called 'batch science', would use laboratory automation to address areas of research that its advocates say are being held back because existing laboratories lack the capacity to collect and analyse the necessary volumes of data.

They argue that existing biology laboratories in universities and government agencies are too small and labour-intensive to produce the experimental data to answer important scientific questions.

Questions suited to the approach include the characterization of pathogens carrying infectious diseases, so as to anticipate the mutation of virulent strains; rapid assessment of agents used in biowarfare or bioterrorism attacks; and the screening of food products to ensure food safety.

Last week in Washington, about 200 scientists from different disciplines met at the National Academy of Sciences to exchange ideas on laboratory automation and batch science. The meeting, "Automation in threat reduction and infectious disease research: need and new directions", was instigated by the University of California at Los Angeles and the Los Alamos National Laboratory in New Mexico. It was also sponsored by the National Academy of Engineering, the Institute of Medicine and three government agencies.

Kumar Patel, a physicist at the University of California and former president of the American Physical Society, who helped to organize the meeting, says the government tends to support traditional models of 'small' or 'big' science, but should be persuaded to build the highly automated, medium-sized laboratories that could do batch science.

While pharmaceutical companies have been using such approaches for drug development, and genome sequencers are engaged in the automated mass production of sequencing data, government agencies and universities analyse samples manually and on a small scale. Agencies such as the Centers for Disease Control and Prevention, which exists to protect the country from infectious disease, have little or no capability for the large-scale processing of samples. Some scientists argue that such a capability, combined with computer analysis of the results, could enable the agencies to function far more effectively.

Patel said people at the meeting "had generally agreed that having such an automated scheme might go a long way toward filling the holes" in these disciplines. **Colin Macilwain**

EMBO seeks European role in E-Biomed

[PARIS] Frank Gannon, executive director of the European Molecular Biology Organization (EMBO) in Heidelberg, will meet Harold Varmus next week to discuss possible European input into E-Biomed.

"Europe has to be intensively involved [in the

discussions]," says Gannon. "I welcome the discussion that has been stimulated by the E-Biomed proposal," says Gannon. "It has brought [debate on electronic publishing] to a head."

Gannon, who shares many of the scientists'

concerns about the proposal, says that "the European Bioinformatics Institute and EMBO are willing to play a part in seeing that there is a successful outcome."

"If it is going to happen, EMBO is uniquely placed to represent European interests."

D. B.

Los Alamos scientist moved secrets to unclassified computer

[WASHINGTON] The US Department of Energy admitted last week that Wen Ho Lee, a Taiwan-born scientist at the Los Alamos National Laboratory, had downloaded some of the most important US nuclear weapons design codes from classified to unclassified computers, where they may have been accessed by unauthorized persons.

The transfer of the codes came to light after the Federal Bureau of Investigation (FBI) finally obtained permission to examine Lee's computer as part of its investigation into his alleged espionage activities at the New Mexico lab. Some commentators say that the movement of the codes could have given China access to many of the United States' most valuable nuclear weapons secrets. It has been reported, but not confirmed, that the downloading took place in 1995, during the first Clinton administration.

The energy department is expected to announce disciplinary action against the officials at Los Alamos who allowed Lee to retain security clearance after the FBI first suspected his involvement in espionage.

The new revelations will lend impetus to draconian legislation, proposed by Senator

Richard Shelby (Republican, Alabama), that would allow congressional committees to vet all visitors to the weapons labs from China, Russia and other sensitive countries.

Industry and greens unite against conservation law

[MONTREAL] Canada's federal government's plans for new endangered species legislation have been criticized by a seemingly unlikely alliance of industry and environmental groups. The position of the self-organized "species at risk working group" — whose members range from the Canadian Pulp and Paper Association to the Sierra Club of Canada — mirrors the one taken by more than 600 scientists (see *Nature* 398, 9; 1999).

These groups, previously at loggerheads on other issues, say the legislation "is likely to fail". They want governments to encourage landowners and others to protect habitats by offering "appropriate compensation". They also say listing decisions should be made by an independent scientific body.

Duma official accused of accepting bribe

[MOSCOW] The head of the staff of the international affairs committee of the Russian Duma, the lower house of

parliament, has been accused of receiving a bribe "for lobbying for the interests of one of the major scientific centres". Vladimir Trofimov was caught by the Federal Security Service receiving US\$5,000, according to a government announcement.

Unofficial reports say that the arrest is being linked with the Joint Institute for Nuclear Research in Dubna, near Moscow. The committee has drawn up new regulations for the administration of this centre, to give it international status and so guarantee it financing and substantial privileges. The proposals were due to be discussed by the Duma on 21 May. It is also rumoured that Trofimov was to have received a further \$5,000 after the regulations were approved by the Duma.

Peres seeks backing for 'peace' research institute

[SAN DIEGO] Former Israeli Prime Minister Shimon Peres has joined the board of the Scripps Research Institute in La Jolla, California. He hopes to make contacts in the US scientific community that will help create a joint Israeli-Palestinian research institute in the Middle East, advancing the peace process.

Peres plans a research institute to form part of his Peres Center for Peace in Tel Aviv, and has reportedly begun raising funds.

“Peres believes that the only way to have peace is to make sure Israelis and Palestinians are equal intellectual partners,” says Scripps president Richard A. Lerner.

Eurohorcs seek study of animal protection laws

[MUNICH] The heads of European research organizations, known collectively as the Eurohorcs, agreed last week to ask the European Science Foundation to prepare a position paper on the status of animal protection laws across the continent.

The Eurohorcs' discussion was led by Ernst-Ludwig Winnacker, president of the German research council, the Deutsche Forschungsgemeinschaft. He is fighting a move to introduce a clause into the German constitution to protect the rights of animals (see *Nature* 397, 461; 1999). Winnacker points out that a recent survey revealed most Germans to be against animal experiments in general, but in favour of experiments that lead directly to health benefits.

US isolated over stand on labelling GM foods

[LONDON] **The United States appears to be losing support for its position on the labelling of genetically modified (GM) foods. At a**

meeting of the United Nations food standards body, the Codex Alimentarius Commission, in Ottawa last week, Argentina alone supported the US view that labels should be printed only on foods with detectable GM ingredients that differ substantially from equivalent non-GM varieties.

Last year, the United States had the support of Brazil, Australia, New Zealand, Canada, Mexico and Peru. Some of these countries are moving closer to the position of the European Union: that all foods with detectable GM ingredients should carry a mandatory label (see *Nature* 398, 641; 1999).

Howard Hughes gives more to eastern Europe

[WASHINGTON] The Howard Hughes Medical Institute announced last week that it will continue to support scientists in Eastern Europe and the former Soviet Union, providing grants worth \$15 million over the next five years. Life scientists from the region are being invited to apply for grants worth \$40,000 to \$90,000 a year, running for five years from November 2000.

The 90 scientists in the region supported by the institute's existing five-year programme, which expires next year, will not have their grants automatically renewed, but are free to reapply, said a spokeswoman. Applications

will be invited from young scientists who already have full-time positions at non-profit institutions, and a publication track record in English-language international journals. More information is available at www.hhmi.org/grants/international.

US immunologists share \$440,000 Japan Prize

[WASHINGTON] Two US scientists who isolated and characterized human histocompatibility proteins have been awarded the Japan Prize, valued at ¥50 million (US\$440,000).

Jack Strominger and Don Wiley, professors of biochemistry at Harvard University, received the prize last week. It recognizes their work over three decades to elucidate the structure and function of major histocompatibility proteins, which are key to the immune system's recognition of foreign substances. It is hoped that the work will lead to vaccines against tumours and infectious diseases, improvements in organ transplantation and better understanding of autoimmune diseases.

The other Japan Prize — two are awarded each year — went to W. Wesley Peterson for his research on digital communications error control. Peterson is a professor of information and computer sciences at the University of Hawaii at Manoa.



The full text of items on this page — and further news and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 1 July — can be found on *Nature's* website at <http://www.nature.com>

Science told to slow down and let ethics debates catch up

[MONTREAL] If scientific discoveries threaten to outstrip efforts to place their application in an ethical context, “science will need to wait and to help ethics to catch up”, according to a leading Canadian bioethicist.

Margaret A. Somerville, a law professor at McGill University's Centre for Medicine, Ethics and Law in Montreal, says that “science time, medical time, business time, political time and ethics time” are on different scales, and that “this can be a source of serious difficulty in doing ethics regarding our new science”.

She points out that scientists “have an enormous drive and enthusiasm for discovery”, and that business also wants to proceed as quickly as possible, while politicians are subject to the ‘do something’ pressure of the electorate, especially in relation to threats to health or life, such as the shortage of organs for transplants.

“In contrast, ‘doing ethics’ in relation to the new science may not be able to be speeded up. There may be an irreducible minimum time needed both to obtain the necessary facts on which to base good ethics, and for a ‘sedimentation-of-values’ process to take place.”

She also suggests that a minimum amount of time is needed for the public to become familiar with the meaning of a scientific development in terms of its potential benefits, risks and harm, “not only at the physical level, but also at the level of its potential impact on values, norms, traditions, customs, cultures, beliefs, attitudes, and so on”.

Somerville argues that there is a difference between simply delivering information on ethics to the public, and engaging the public in discussion about a new technology. “The latter takes time. Indeed, how to engage the public adequately and effectively in ‘ethics talk’ is a difficult question.”

The instant worldwide access to scientific developments also raises ethical issues, she argues. For example, countries that lack the research infrastructure needed to develop xenotransplantation can use this technology when it has been developed elsewhere.

“But they might not do so safely for either the research subject/patients involved, or the public, and they may not do so ethically. In short, many new scientific techniques, such as xenotransplantation, need universal ethical regulation.”

Full text: <http://helix.nature.com/wcs/c13.html>

Africa ‘needs a linguistic and gender revolution’

[HAMMAMET, TUNISIA] An African renaissance is unlikely to happen while English remains the main medium for science education, and without concerted efforts to involve women, according to Ali Mazrui, professor emeritus of Africana Studies at Columbia University, New York.

In a keynote address to the fifth general conference of the African Academy of Science, in Hammamet, Mazrui said that one reason why science had failed in Africa was the “masculine bias” of educational institutions both before and after independence.

Another reason, he suggested, was the failure of both the colonial powers and the continent's independence leadership not only to acknowledge indigenous African medicine and technology, but also to convey discoveries and ideas in modern science in Africa's indigenous languages.

“Africa's educational system is good at transmitting the Western literary heritage, but not the scientific heritage,” said Mazrui. He pointed out that Julius Nyerere, first president of Tanzania, translated Shakespeare's *The Merchant of Venice* and *Julius Caesar* into kiswahili. “Why couldn't anyone do the same for Darwin's *On The Origin of Species*, or *The Descent of Man*?” he asked.

Similarly Milton Obote, a former president of Uganda, had taken his first name



Mazrui: shake off the imperial legacy.

from the poet John Milton. “Why has there not been an African Newton Obote, out of admiration for the great physicist?”

The provision of science education in indigenous languages must be a priority for any future African development policy, said Mazrui. Countries such as Japan and South Korea had developed using their own languages. “Can Africa take off while it is hostage to the languages of former imperial powers?” he asked.

“When Japanese physicists meet, they can discuss advanced physics in Japanese,” said Mazrui. “But, when two African economists meet, they are unable to use any language but English, even if they come from the same African linguistic group.”

Mazrui added that Africa was unlikely to develop unless the role of women was recognized as being on a par with that of men. “In African societies, women are the trustees of indigenous languages. A linguistic revolution needs to be reinforced by a gender revolution — an androgynization of science.”

Full text: <http://helix.nature.com/wcs/c14.html>

Unesco pushing for a ‘Decade for Science’

[LONDON] Unesco officials may ask the United Nations (UN) general assembly to declare the first decade of the millennium the ‘Decade for Science’, as a follow-up to the forthcoming World Conference on Science.

The idea has obvious appeal as a way of boosting follow-up to the conference. But support for such a declaration will not be easy to obtain, as the UN has already named that period the international decade for the culture of peace and nonviolence for the children of the world. Similarly, 1997–2006 is the decade for the eradication of poverty, 1995–2004 is the decade for human rights education, and 1995–2004 is the decade of indigenous people.

A second option would be to declare a single year the ‘year of science’. But the UN calendar is full until 2006. Peace and thanksgiving will be observed in 2000, with volunteers, mountains, racism and xenophobia, ecotourism and ‘small loans’

pencilled in for the following years.

David Wardrop, honorary secretary of the UK–Unesco Forum, says that the idea of parallel themes is becoming less popular with governments, as multiple themes make less of a global impact than one. The current decade, for example, has no fewer than five separate themes: eradication of colonialism; development; natural disaster reduction; drug abuse; and international law.

UN annual and decadal themes are designed primarily to raise public and political awareness on social and environmental issues. Wardrop says that decadal themes are more likely to lead to practical policies than single years devoted to issues, as there is more time for governments and non-governmental organizations to mobilize action. The decade for the disabled, for example, is widely considered to have been a success.

Full text: <http://helix.nature.com/wcs/a29.html>

Editors' responsibility in defeating fraud

Sir — As editor-in-chief of *Nutrition*, an international medical journal, and as director of a research laboratory, I found your Briefing on science and fraud most interesting, because I am both a producer and a consumer of science (*Nature* **398**, 13–17; 1999).

My editorial colleagues and I have a high state of awareness of 'fabrication, falsification and plagiarism (FFP)'. As reviewers of manuscripts, we have a difficult time detecting the two Fs, but allegations of the P have come to our attention several times.

I believe that editors have an obligation to the scientific community to pass such concerns to the authors and to their institutes' research dean or administrative supervisor in a confidential manner for investigation according to the guidelines of

the US Office of Research Integrity. In doing so, we do not act as "secret police", as the editor of the *Journal of the Norwegian Medical Association* maintains. Instead, we align ourselves with the UK Committee on Publication Ethics and the World Association of Medical Editors, whose recommendations are in my view appropriate.

It does untold harm to the scientific community to be betrayed, deceived and defrauded. Such harm ranges from the squandering of limited research resources to the undermining of confidence and trust in the reporting of scientific findings. A journal should not be used to validate misconduct by publishing fraudulent data submitted knowingly by the author. If this occurs, editors bear an obligation to retract the paper.

Our journal asks authors to sign a declaration of scientific integrity in their letter of transmittal.

To avoid scientific misconduct in my laboratory, each new research fellow's attention is drawn to this potential problem via policy and procedure material given to them on arrival, and the consequences of such temptations are clearly spelled out. Each new fellow also repeats a portion of their predecessor's work to confirm the results, as an internal control standard. This has not dampened the lust for data among the 'young and hungry'. But, ultimately, solid, reliable laboratory habits and supervision and mentoring are critical components to prevent misconduct.

Michael M. Meguid

*Nutrition, Department of Surgery,
750 E. Adams St., Syracuse, New York 13210, USA*

Inequities in German research system

Sir — You report that the German government, which has been making efforts to rebuild the science base in east Germany, is changing its funding policy, because research in the east is now considered ready to compete with the west (*Nature* **398**, 7; 1999). We would like to add a point that is often forgotten: salaries in east German government-funded research labs are generally only at 86 per cent of the level of their west German counterparts.

This difference was introduced after the country's reunification to take account of the former lower productivity of companies, administrations and labs in the east. Curious effects can be observed as a result. In Berlin, for example, moving to a job across the street can bounce you from 100 per cent to 86 per cent payment (city versus federal funding). This makes it a tough job to hire scientists in the east.

We wonder whether the new east-west competition for federal research funds will in this respect be on a fair and equal footing, and hope that the government will find a way to solve this problem.

Stefan Jähnichen, Klaus-Robert Müller

*GMD FIRST, Rudower Chaussee 5,
12489 (East) Berlin, Germany*

Sir — You report Germany's junior minister for research as saying that east German researchers have been so well supported in recent years that they are now considered good enough to compete with west German scientists for conventional

funds. Having worked in research institutes in west Germany, Britain and now in east Germany, I can assure you that it is not the standard of the scientists that has been the limiting factor, but the administration with which they have had to contend.

We have good and bad scientists, like everywhere else, and it says much for the spirit of the good ones that they show every sign of succeeding in the face of adverse conditions and patronizing attitudes.

Michael Cross

*Interdisciplinary Centre for Clinical Research,
University of Leipzig,
Johannisallee 30a, 04103 Leipzig, Germany*

Cloning claims challenged

Sir — Karl Illmensee¹ has suggested that the data on mouse cloning he and Hoppe² published in 1981 were unjustly neglected or rejected. He opines that, following the recent cloning of mice³, "the time has come for the correct evaluation of our earlier results on the first cloning of a mammal" [my italics].

Illmensee and Hoppe² reported the birth of live mice following the transfer of nuclei from inner cell mass (ICM), but not trophoblast, cells into an enucleated zygote. Their method involved denuding the donor nucleus, injection of the donor nucleus into the zygote cytoplasm, and using the same pipette to remove the zygotic pronuclei. To my knowledge, neither this method nor these results were ever repeated⁴, though not for want of trying.

Tsunoda and Kato³ recently reported the birth of live mice following the transfer of ICM or trophoblast nuclei into an enucleated oocyte. Their method was as follows: metaphase chromosomes were removed without penetrating the egg membrane; a single ICM or trophoblast cell was introduced under the zona pellucida and fused with an enucleated oocyte using inactivated Sendai virus; and the nuclei of reconstituted eggs that developed to the two-cell stage were fused, again using Sendai virus, with enucleated blastomeres of normal two-cell stage embryos.

We described this method in 1983 (ref. 5), and it has been used by everybody doing nuclear transfer ever since. We do not know all of the parameters necessary for successful nuclear transfer, but it seems that oocytes are much better recipients than zygotes, and it may be that cloning cannot be achieved by the transfer of somatic nuclei into zygotes.

Comparison of Illmensee and Hoppe's methods and results² with those of Tsunoda and Kato³ does not show that their early data should be re-evaluated and rescued from neglect. One could equally well argue that the success of the Apollo missions confirms Jules Verne's or Cyrano de Bergerac's descriptions of voyages to the Moon.

Davor Solter

*Max-Planck Institute of Immunobiology,
Stuebeweg 51,
79108 Freiburg, Germany*

1. Illmensee, K. *Nature* **398**, 19–20 (1999).

2. Illmensee, K. & Hoppe, P. *Cell* **23**, 9–18 (1981).

3. Tsunoda, Y. & Kato, Y. *J. Reprod. Fertil.* **113**, 181–184 (1998).

4. McLaren, A. *Nature* **309**, 671–672 (1984).

5. McGrath, J. & Solter, D. *Science* **220**, 1300–1302 (1983).

From youthful PMs to presidents who walk tall

Sir — M. de L. Brooke reasons that British and US voters will elect a leader from an age group that they respect. On the basis of an analysis of the age at which British prime ministers first came to power, he concludes that “voters give the job to somebody whose attributes are honest signals of ability in the society of the day” (*Nature* 398, 102; 1999). Both reasoning and conclusion are flawed.

First, Brooke fails to ask whether the ages of victorious prime ministers differ significantly from those of their vanquished opponents. The two-party nature of the British and US electoral systems makes this a vital question. If they do not differ significantly, then voters *cannot* choose on the basis of age — the choice has already been made for them. If they do differ, then we need to know in what direction(s).

Second, it is at least questionable whether people vote for prime ministers at all. There is plenty of psephological evidence to show that it is the political party that commands people's loyalty. This can be demonstrated anecdotally. Winston Churchill was an immensely popular war leader, but he was heavily defeated in the 1945 general election. Jim Callaghan was more popular than Margaret Thatcher, but he lost in 1979. It is also significant that Brooke does not find an age trend in the US system, where there is a more direct relationship between voter and president.

It might be possible to rescue at least a version of Brooke's thesis by shifting the emphasis away from voter choice to party structure. That is, to claim that the reason that prime ministers have become progressively younger in the twentieth century is that there is a perception on the part of the major players in political parties that the voting public has less respect for older people than in the past. But there are still problems. First, the analysis would no longer be founded on a notion of ‘honest signals of quality’. And second, it would be to oversimplify the process whereby an élite rises to the top of a party. For example, Margaret Thatcher — in electoral terms, the most successful prime minister of this century — became leader of her party in 1975 in circumstances that had nothing to do with her age (and a lot to do with Sir Keith Joseph's committing political suicide by a speech in Birmingham; and Edward du Cann's withdrawing from the Conservative leadership contest at the last moment). And Tony Blair became leader of the Labour party only after the sudden death of John Smith.

As a philosopher, I frequently try to persuade colleagues of the merits of a

sophisticated reductionism. Unfortunately, in this instance, the reductionism is not sophisticated and the conclusions are, therefore, likely to be erroneous.

Incidentally, I wonder if your correspondent is aware that every US president elected between 1900 and 1968 was taller than his major opponent.

Jeremy Stangroom

The Philosophers' Magazine,
98 Mulgrave Road, Sutton, Surrey SM2 6LZ, UK

Reaping the benefits of cropping experiments

Sir — There is no doubt that intensive agriculture has significant environmental impacts. There is also a widespread belief that ‘conventional’ agriculture is the problem and ‘ecological’ methods the solution. David Tilman¹ discusses a single experiment² that compared legume-based and farmyard manure-based cropping systems with a fertilizer-based one. Can such experiments be used as the basis for general statements about agriculture?

Comparisons of cropping systems often have more of an illustrative than an explanatory value, since there are so many factors differing. There is a risk that the conclusions that can be drawn from comparisons between very different systems are either self-evident or obscure.

In this case, two carbon-exporting systems are compared with a carbon-recycling one (manure). Soils in all treatments received equal amounts of carbon input. But manure input to soil is a more stable carbon source than fresh plant residues. Manure is produced by feeding, for example cattle, and the cattle gain energy from the more accessible carbohydrates, leaving more recalcitrant compounds in the manure. Therefore, it should be no surprise that soil carbon levels over a 15-year period increase more from adding recalcitrant (manure) carbon than from adding carbon in fresh crop residues.

The increase in soil carbon owing to legumes in the crop rotation may partly be due to differences in the quality of crop residues, but may also be due to differences in longevity of the crop, higher water uptake (drier soil gives lower carbon losses) or differences in soil cultivation. Usually, legume or grass leys increase soil carbon compared with annual crops, owing to these factors.

When looking at cumulative nitrate leaching, there was a 50% higher rate from the fertilizer-based system, although not statistically significant ($P = 0.06$). But the whole difference for the 15-year period is based on the observation that only one year

out of the five measured had a higher leaching rate. This is interesting, but there is no attempt to discuss this difference.

Much valuable information can be obtained by comparing different cropping systems in long-term experiments.

However, as long as the interpretation is biased and the information provided to the reader is incomplete, conclusions will remain more ideological than scientific.

Olof Andrén, Holger Kirchmann, Olle Pettersson

Department of Soil Sciences,
Swedish University of Agricultural Sciences,
Box 7014, S-75007 Uppsala, Sweden

Tilman replies — Andrén and co-authors express scepticism about the possibility that ‘ecological’ methods may help solve agricultural pollution², but do not dispute the equality of yields from manured versus fertilized crops. I highlighted¹ experimental work on ‘ecological’ farming² not because it offered a definitive solution to agricultural pollution, which it did not do, but because it demonstrated the plausibility of approaches that have been ignored in most conventional research. Any practices that seemingly reduce pollution while maintaining yields and profitability are worthy of study.

Agricultural intensification during the past 35 years has led to a doubling of world grain production, but this required 6.9- and 3.5-fold increases in annual global rates of nitrogen and phosphorus fertilization, respectively, and a doubling of irrigated land³. Use of pesticides also increased markedly, with many of these accumulating far from points of application⁴. If these trends presage the future, the next doubling of global food production, expected within 35 years, will require a tripling of nitrogen and phosphorus fertilization and a doubling of irrigation³.

Nitrogen and phosphorus losses from agricultural fields are already the major source of nutrient loading into freshwater and nearshore marine ecosystems⁵, and are a major source of terrestrial nutrient loading⁶. This pollution is having serious impacts on non-agricultural ecosystems⁵⁻⁶.

Research is needed that pursues all reasonable approaches to this problem. The apparent distrust between conventional and ‘ecological’ schools of agricultural thought must not blind either side to novel insights, nor slow the development of solutions to a global problem.

David Tilman

Department of Ecology, Evolution and Behavior,
University of Minnesota,
St. Paul, Minnesota 55108, USA

1. Tilman, D. *Nature* 396, 211–212 (1998).

2. Drinkwater, L. E. *et al.* *Nature* 396, 262–265 (1998).

3. Tilman, D. *Proc. Natl Acad. Sci. USA* 96 (in the press, 1999).

4. Blais, J. M. *et al.* *Nature* 395, 585–588 (1998).

5. Carpenter, S. *et al.* *Ecol. Applications* 8, 559–568 (1998).

6. Vitousek, P. *et al.* *Ecol. Applications* 7, 737–750 (1997).

Diamondoids and oil are not forever

Martin Schoell and Robert M. K. Carlson

A measure of the amount of oil destroyed by high temperatures deep in the Earth may help to gauge the depths to which oil extraction remains commercially viable, and so be useful in estimating ultimate oil reserves.

Hydrocarbon-based fuels make the world go round, and will do so until we run out of the crude oils from which these energy-rich liquids are derived. It may not be a pressing issue in these times of abundant oil and low prices, but it is clear to all involved in oil exploration that this precious resource is limited. Exactly when crude oil will run out depends in part on the depths at which it can be found in the Earth. Its preservation in the Earth's crust is limited, because at temperatures above about 200 °C most crude oil components start to decompose, or 'crack', to smaller and smaller compounds, and ultimately to methane gas and graphite. According to a paper by Dahl *et al.* on page 54 of this issue¹, the extent of oil destruction could be assessed with a group of especially resilient, naturally occurring hydrocarbon compounds that resemble diamonds in their chemical structure, and which chemists call 'diamondoids'.

Natural crude oils form over millions of years under the action of heat deep within the Earth. Petroleum geochemists have coined the phrase 'oil window' for the temperature range within which oil is liberated from organic-carbon-rich rocks in the Earth's crust. The prevailing window gives an upper temperature limit for oil preservation, because most hydrocarbons cannot tolerate extreme heat over extended geological times.

It is not clear what this temperature limit is, and what geological time spans are required to crack oil. This issue has not yet been resolved, in part because of a lack of quantitative measurements that can relate oil composition to the extent of cracking. For instance, because oil is converted to gas, in principle the gas-to-oil ratio should increase in a high-temperature environment. However, the reservoirs in which oil and gas are contained are not gastight, and leaking gas makes this measurement a questionable solution for our problem.

Enter the resilient diamondoids: this hydrocarbon family begins with adamantane, a molecule with ten carbons ($C_{10}H_{16}$). Adamantane (from the Greek name for diamond) has a volatility similar to that of jet fuel but a melting point above that of metallic tin. Higher members of the family have four additional carbons added as a three-dimensional cage (Fig. 1), so we have diamantane ($C_{14}H_{20}$), triamantane ($C_{18}H_{24}$) and so on (including alkyl-substituted analogues). Where is the line drawn between these higher diamondoids and 'true' diamonds? For diamonds with surfaces that terminate in hydrogen, there really is no such line, and diamondoids could be viewed as tiny diamonds, adamantane being the smallest possible example, with a carat size of about 10^{-21} (a carat is equal to one-fifth of a gram).

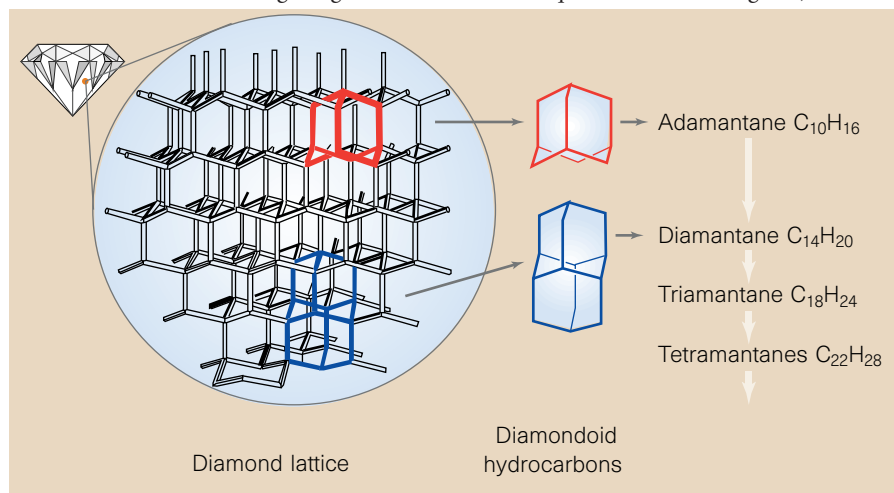


Figure 1 Chemical structures of diamond-like molecules, or diamondoids, found in petroleum. Dahl *et al.*¹ show that diamondoid concentrations in oil can be used to estimate the extent of oil destruction in a particular reservoir.

It is easy to make adamantane and its methylated analogues in the laboratory from other hydrocarbons², but the higher molecular-weight diamondoids are extremely difficult to synthesize³ (to date, the largest is a tetramantane, $C_{22}H_{28}$). But tetramantanes, pentamantanes and even hexamantanes occur naturally in some deep gas deposits^{4,5}. As natural gas is produced from deep in the Gulf of Mexico and elsewhere, literally tonnes of diamondoids can condense as crystalline solids in the flow lines (Fig. 2), compromising safety and causing costly production shutdowns.

Ever since diamondoids were first identified in crude oils in 1933 by Landa and Machacek^{6,7}, geochemists have wondered about the origin and fate of these compounds in the Earth. Diamondoids are not biological products. So when, where and how are they synthesized in nature? Why do they become more abundant in deep, hot, geological environments? The secret of the resilience of diamondoids is that their diamond-like chemical structure is less prone to thermal destruction than other hydrocarbons in the Earth's crust. The chemical stability of diamondoids is well known, and they have been used as internal standards for coal liquefaction studies⁸ and as a matrix for free-radical reactions⁹ (because they are not destroyed by these reactions). Similarly, in the struggle for survival at high temperatures in the Earth, diamondoids outlive other hydrocarbons found in crude oils.

How can the relative stability of diamondoids be used to assess the ultimate depth of crude-oil preservation? Dahl *et al.*¹ demonstrate in laboratory experiments that diamondoid concentrations increase in a predictable fashion with increasing temperature. This opens the possibility of using diamondoid concentrations as a measure of oil destruction. Using the Gulf of Mexico as a natural laboratory, they show that oils from deep reservoirs have undergone considerable destruction and are up to 80% cracked to gas. From this we can conclude that the oil



Figure 2 Diamondoids collected from a gas field in the Gulf of Mexico in which they have accumulated. The vial is about 2.5 cm in diameter.

resources in the Gulf of Mexico are limited to strata shallower than the deep Jurassic (5–6 km deep). With this new chemical measuring stick we can now — in principle — screen other sedimentary basins for the ‘floor’ of their hydrocarbon production, and make more realistic estimates of the world’s oil reserves.

Many questions remain regarding these fascinating molecules. Where do the diamondoids come from — particularly the larger, hard-to-synthesize ones — and is their concentration in unaltered oils always the same? This is the biggest problem facing their use as quantitative indicators of oil cracking. Their concentrations in crude oils vary for unknown reasons, possibly related to the types of minerals in oil source rocks or reservoirs. Another nagging question is whether diamondoids themselves are eventually destroyed in high-temperature environments. We know from laboratory experiments¹⁰ that they can be destroyed at temperatures of 550 °C, and carbon-isotope analyses of diamondoids from the Gulf of Mexico¹¹ show that some of them are already affected by cracking processes. Evidently, then, diamondoids are not forever. Like all hydrocarbons they are perishable, although they appear to survive at sufficiently high temperatures to serve as good chemical tracers as oils are increasingly converted to gas and a carbonaceous residue called pyrobitumen, deep in the Earth’s crust.

Will this new measuring stick finally end a long-lasting debate? Despite a large body of experimental evidence on oil cracking¹², there are surprising reports that oils may survive very high temperatures¹³. High pressure may also slow the destruction of oil¹⁴, suggesting that oil might be preserved under special circumstances in deep sedimentary basins. Should we re-examine our models of oil stability in deep basins, and does this mean that there is much more oil than previously thought? These issues could not be resolved until now because quantitative measurements of the extent of oil cracking in natural reservoirs were not available. Maybe, with the help of these 10^{−21}-carat diamondoids, petroleum scientists will eventually be able to tell us how much oil is left and how long we can enjoy an easy life with abundant energy and low fuel prices. □

Martin Schoell and Robert M. K. Carlson are at the Chevron Research and Technology Company, 100 Chevron Way, PO Box 1627, Richmond, California 94802, USA.
e-mails: mrtm@chevron.com
crrm@chevron.com

1. Dahl, J. E. *et al.* *Nature* **399**, 54–57 (1999).
2. Fort, R. C. & Schleyer, P. V. R. *Chem. Rev.* **64**, 277–300 (1964).
3. McKervy, M. A. *Tetrahedron* **36**, 971–992 (1980).
4. Lin, R. & Wilk, Z. A. *Fuel* **74**, 1512–1521 (1995).
5. Tominaga, K. & Haga, M. *Chem. Econ. Eng. Rev.* **17**, 23–30 (1985).
6. Landa, S. & Machacek, V. *Collection Czech. Chem. Commun.* **5**, 1–5 (1933).
7. Petrov, A. *et al.* in *Advances in Organic Geochemistry 1973* (eds

Tissot, B. & Biennier, F.) 517–522 (Editions Technip, Paris, 1974).

8. Aczel, T., Gorbaty, M. L., Maa, P. S. & Schlosberg, H. *Fuel* **58**, 228–230 (1979).

9. Tegowski, A. T. & Pratt, C. W. *J. Am. Chem. Soc.* **106**, 64–67 (1984).

10. Oya, A., Nakamura, H., Otani, S. & Marsh, H. *Fuel* **60**, 667–669 (1981).

11. Schoell, M. *et al.* *18th International Meeting on Organic Geochemistry Abstr.* 55–56 (Forschungszentrum Juelich, Juelich, 1997).

12. Behar, F. *et al.* *Revue de l’Institut Français du Pétrole* **46**, 151–183 (1991).

13. Price, L. C. *Geochim. Cosmochim. Acta* **57**, 3260–3281 (1993).

14. Hill, R. J. *et al.* *Energy and Fuel* **10**, 873–882 (1996).

Physiology

A calcium window to the gut

G. David S. Hirst

Gut movements occur in such an ordered way that the sophistication of the system that controls them is often overlooked. Gut contractility is determined by the internal concentration of calcium ions in the smooth muscle cells that line the gut. This, in turn, is controlled by a balance between the influences of neuronal circuits in the gut wall, and those of non-neuronal ‘pacemaker’ cells. On page 62 of this issue, Stevens *et al.*¹ report how, using a new approach, they have monitored changes in the internal concentration of Ca²⁺ over whole segments of the colon during natural gut activity. They find that the local nervous system first initiates Ca²⁺ waves, which move from cell to cell in the muscle layer, and then directs them along the gut wall in an ordered way.

The gut is a hollow tube. Its wall is made up of two layers of smooth muscle cells — an outer longitudinal layer, which contracts to shorten the gut, and an inner circular layer, which undergoes sphincter-like contractions. Between these two muscle layers lies an extensive network of pacemaker cells, which are neither nerve nor muscle cells. Known as the interstitial cells of Cajal (ICC) these cells were identified many years ago, yet their importance in controlling gut motility has only recently been recognized². The ICC generate repeating, large-amplitude waves of depolarization, which spread to the two muscle layers, triggering coordinated

contractions³ (Fig. 1). These ‘myogenic’ rhythms, initially generated by the ICC, cause a continuous pattern of regular local contractions, often termed segmentation.

Neurons are also involved in controlling gut movements, allowing food to be broken down, nutrients to be absorbed and waste products to be expelled. Indeed, there are as many neurons in the gut, controlling the movements there, as in the spinal cord to control most other body functions. The intrinsic neurons mediate complete local reflexes, and there are populations of neurons with sensory, motor and interneuronal functions. They form two major nerve plexuses, or networks, which run from the start to the end of the gastrointestinal tract⁴. One plexus, the myenteric, lies between the two muscle layers. Motor neurons innervate the muscle layers and, depending on the chemical transmitter that they release, these neurons either excite or inhibit the muscle⁵. The sensory neurons monitor the nature and amount of the gut contents, and the interneurons relay information from neuron to neuron down the gut. The second nerve net, the submucous plexus, lies between the circular muscle layer and the overlying mucosal membrane. Together these plexuses support an intrinsic reflex — the peristaltic reflex — that moves the gut contents onwards.

How do the separate muscular and neural elements work together? New insight now

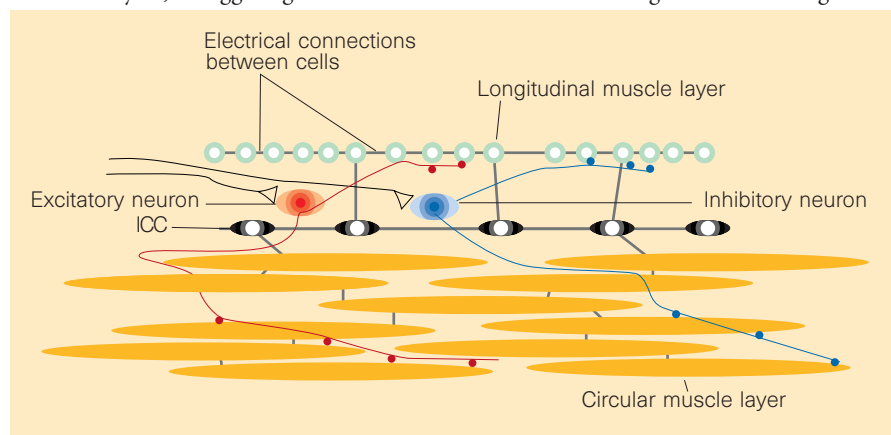


Figure 1 Two-way control of the gut. Smooth muscles cause gut movements under the command of a neuronal influence (inhibitory and excitatory myenteric neurons), and a non-neuronal influence (the interstitial cells of Cajal; ICC).

comes from Stevens *et al.*¹. The authors loaded all the muscle cells in intact guinea pig bowel preparations with a fluorescent dye, to measure the internal concentration of Ca^{2+} . On a cellular scale they detected discrete Ca^{2+} waves, which spread through the longitudinal muscle layer. By tracking the sites of origin and spread of the internal Ca^{2+} waves, Stevens *et al.* could describe their behaviour. They found that Ca^{2+} waves in the longitudinal layer result from neuronal activity, because these waves can be suppressed by drugs that block the effects of neurally released excitatory transmitter. The waves spread readily along the axis of the colon, but poorly around the circumference. Such uneven, or anisotropic, conduction reflects the ease with which ionic currents spread along and across the cells that make up this longitudinal layer.

Stevens and colleagues also found that the firing of inhibitory nerves produces local barriers of inhibition, which limit the spread of the excitatory wave and allow an increase in the internal concentration of Ca^{2+} . This, in turn, triggers local contractions, which mix the contents of the gut. When this peristaltic reflex is activated, descending inhibitory waves abolish the discharge of Ca^{2+} waves, allowing the region to relax. After the period of inhibition, Ca^{2+} waves are again detected and the ensuing contraction moves the gut contents onward in an ordered way. So, the intrinsic nervous system acts not only to initiate localized movements, but also as a com-

mand system that oversees the movement of gut contents over longer distances.

An exciting aspect of Stevens and colleagues' work is that it brings together information obtained from a molecular approach with a more integrated physiological problem. Over the past decade there have been dazzling advances in our knowledge of membrane channels, intracellular messengers, control of the internal Ca^{2+} concentration and genetic profiles. But many of these advances have overlooked the fact that we often do not know what the elements involved do in the body. It will be very productive to apply the knowledge base from more reductionist approaches to physiological problems. One example is the use of gene manipulation to identify the role of the ICC in gastrointestinal motility⁶. Measuring changes in the internal concentration of Ca^{2+} during peristalsis is another step in this direction. □

G. David S. Hirst is in the Department of Zoology, University of Melbourne, Parkville, Victoria 3052, Australia.

e-mail: georgedh@myriad.its.unimelb.edu.au

1. Stevens, R. J., Publicover, N. G. & Smith, T. K. *Nature* **399**, 62–66 (1999).
2. Sanders, K. M. *Gastroenterology* **111**, 492–515 (1996).
3. Dickens, E. J., Hirst, G. D. S. & Tomita, T. *J. Physiol.* **514**, 515–531 (1999).
4. Furness, J. B. & Costa, M. *The Enteric Nervous System* (Churchill-Livingstone, Edinburgh, 1987).
5. Burnstock, G., Campbell, G. D. & Rang, M. J. *J. Physiol.* **182**, 504–526 (1966).
6. Ward, S. M., Burns, A. J., Torihashi, S. & Sanders, K. M. *J. Physiol.* **480**, 91–97 (1994).

Mesoscopic vortex physics

From vortices to genomics

Terence Hwa

A basic property of superconductors, besides their lack of electrical resistance, is the ability to exclude magnetic fields from their interior so long as the applied field does not exceed a critical value H_c . In the case of type-II superconductors, which include the high- T_c materials, screening of external fields is only partial, with magnetic flux penetrating the interior in the form of vortex lines below a lower critical field H_{c1} . Previous research in this area has focused on the collective behaviour of large numbers of vortex lines. Now, on page 43 of this issue, Bolle *et al.*¹ report observations of the kinetics and thermodynamics of individual vortex lines threaded through a planar type-II superconductor. A series of fascinating discrete 'events' is observed close to the lower critical field H_{c1} , where a small number of vortex lines are present in the sample. The findings are comparable to 'mesoscopic' electronic systems (typically a micrometre or less in size), where interesting effects such as Coulomb blockade and universal conductance fluctuation occur

when the quantization of electronic charge becomes important in a small volume.

The events observed by Bolle *et al.* are sharp responses to small changes in the ambient conditions, such as the applied field and temperature. Amazingly, the noisy-looking response curves are mostly reproducible (see their Fig. 2 on page 44). These intermittent, staircase-like curves depend on the details of the sample under study, and can be regarded as its fingerprint. The intermittent responses reflect an intricate balance of the different forces acting on the vortex lines: the attraction of the vortex lines to the material defects that are distributed randomly in the sample; a 'line tension', which aligns the average orientation of the vortex lines to the applied field; and the magnetic repulsion, which prevents different vortex lines from going to the same defects.

An underlying physical picture has emerged from detailed theoretical studies referred to by Bolle *et al.*: in equilibrium, the vortex lines find a special configuration of trajectories that minimizes the overall

free energy for the specific arrangement of random defects in the sample. In such a minimal state, a small change in the ambient conditions is not likely to produce any significant response by the already-optimized vortex configuration. However, when the system is forced to change, for example when an extra vortex line is added, a major rearrangement of vortex trajectories takes place, leading to the intermittent, large responses observed in the experiment. This phenomenon can be thought of as the equilibrium analogue of avalanches studied in the context of self-organized criticality (for a review, see ref. 2).

Much of the above picture has been speculated qualitatively in one form or another in studies of the spin-glass model, which has been taken as a paradigm of disorder-dominated systems in the past 25 years³. However, it is only for vortex systems that quantitative theoretical studies have become possible. Such studies are performed in the context of a class of models known as 'directed polymers'. Figure 1 (overleaf) shows an example of a single directed polymer in a '1+1' dimensional random medium. The term '1+1' (as opposed to '2') is used to emphasize the fact that the directions r and z that span the embedding space of the polymer are not equivalent, as the polymer is oriented on average in the z direction. Theorists have also used the '2+1' dimensional model to describe a vortex line in a three-dimensional superconductor, and the '1+2' dimensional model to describe the interface separating two bulk phases of a three-dimensional ferromagnet.

The directed polymer in a '1+1' dimensional random medium is a rare case of a disorder-dominated system whose statistical properties can be solved in great detail analytically, far beyond the mean-field analysis to which most other systems (including the spin-glasses) are limited. Also, the thermodynamics of individual samples can be computed much more quickly using a transfer-matrix method, when compared with the exponentially long computational time needed for spin-glass-type problems. The combined analytical and numerical approach has yielded a great deal of information concerning the exotic properties of randomly pinned directed polymers. Many of these properties are believed to be generic to disorder-dominated systems. This development is analogous to how exact solutions of one-dimensional quantum systems have shaped our understanding of strongly correlated, many-body quantum systems in general.

Despite the theoretical progress made, experimental studies have been hampered by problems with probing the fluctuations of individual lines in a macroscopic sample. Bolle *et al.*¹ overcame this difficulty by

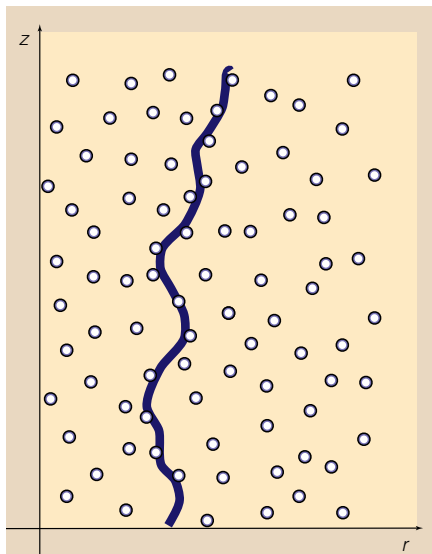


Figure 1 A directed polymer (solid line) in a '1+1' dimensional random medium. The polymer dimensions are described as '1+1' (rather than the usual '2') because the r and z directions that span the embedding space of the polymer are not equivalent, as the polymer is oriented on average in the z direction. The circles represent random disorders that the directed polymer is attracted to. This model has been used to describe the behaviour of individual vortex lines in a planar superconductor, such as that studied in the experiment by Bolle *et al.*¹. The directed polymer model has also been used in a variety of other contexts, ranging from non-Hermitian quantum mechanics to sequence matching in genomics.

attaching a small sample (1.6- μm thick) to a novel micro-mechanical oscillator. The magnetic response of the system was obtained by monitoring small changes in the resonance frequency of the oscillator due to the few vortex lines trapped in the sample. Interaction of these few lines was then used to deduce the properties of a single line. This work is the first quantitative experimental study of this fundamental and fascinating problem of statistical physics. The results obtained are consistent with theoretical predictions, although more detailed experiments are still necessary to make point-to-point comparisons.

The special feature of the directed polymer problem that makes it analytically tractable is the transfer-matrix approach, which expresses the free energy of the polymer recursively in terms of the free energy of a shorter polymer. Huse *et al.*⁴ observed that the 'evolution' of the free energy under successive application of the transfer matrix is equivalent to the dynamics of a randomly forced Burgers' equation, developed in the mid-1970s⁵ in an attempt to understand the hydrodynamics of strongly driven fluids. Much of the knowledge of '1+1'-dimensional directed polymers is derived from exact results of the corresponding Burgers' equation.

The surprising connection between the directed polymer and the hydrodynamics of the Burgers' equation is just the beginning of a long list of problems the directed polymer is related to. Best known is the kinetic roughening of driven interfaces⁶. Other examples include the stick-slip motion between two sliding surfaces⁷, the complex eigenvalue distribution in non-Hermitian quantum mechanics⁸, and — perhaps most surprisingly — the relation⁹ to a 'sequence-alignment' algorithm, which is routinely used by molecular biologists to search for homology between genes and proteins¹⁰. In fact, the sequence alignment problem was formulated as a minimal directed-path problem in the bioinformatics literature as early as 1970¹¹. The transfer-matrix method was proposed back then as a means to find the optimal trajectory of the directed path on a special '1+1' dimensional lattice, with a 'random potential' derived from the composition of sequences being aligned. Today, special-purpose computers and chips have been constructed to satisfy the high demand for such transfer-matrix calculations.

The insight gained from understanding the statistical physics of the directed polymer has already led to very efficient ways of

computing the statistical significance of alignment results¹², the lack of which has limited applications of these powerful tools. It is conceivable that the experimental investigation of mesoscopic vortex physics initiated by Bolle *et al.* may one day lead to ultra-fast physical devices capable of aligning a vast number of sequences in parallel. □

Terence Hwa, currently on leave at the Rockefeller University is in the Physics Department, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0319, USA.

e-mail: hwa@matise.ucsd.edu

1. Bolle, C. A. *et al.* *Nature* **399**, 43–46 (1999).
2. Paczuski, M., Maslov, S. & Bak, P. *Phys. Rev. E* **53**, 414–443 (1996).
3. Binder, K. & Young, A. P. *Rev. Mod. Phys.* **86**, 801–876 (1986).
4. Huse, D. A., Henley, C. L. & Fisher, D. S. *Phys. Rev. Lett.* **55**, 2924 (1985).
5. Forster, D., Nelson, D. R. & Stephen, M. J. *Phys. Rev. A* **16**, 732–749 (1977).
6. Kardar, M., Parisi, G. & Zhang, Y.-C. *Phys. Rev. Lett.* **56**, 889–892 (1986).
7. Cule, D. & Hwa, T. *Phys. Rev. B* **57**, 8235–8253 (1998).
8. Nelson, D. R. & Shnerb, N. M. *Phys. Rev. E* **58**, 1383–1403 (1998).
9. Hwa, T. & Lässig, M. *Phys. Rev. Lett.* **76**, 2591–2595 (1996).
10. Waterman, M. S. *Introduction to Computational Biology* (Chapman & Hall, New York, 1994).
11. Needleman, S. B. & Wunsch, C. D. *J. Mol. Biol.* **48**, 443–453 (1970).
12. Olsen, R., Bundschuh, R. & Hwa, T. in *Proc. 7th Int. Conf. Intelligent Systems for Mol. Biol.* (eds Lengauer, T. *et al.*) (in the press).

Neurobiology

A spine to remember

Per Andersen

How can we remember some information for many years, even after only a short experience of it? On page 66 of this issue, Engert and Bonhoeffer¹ present evidence that when nerve cells receive an intense stream of impulses, they may form new structural and, perhaps, functional connections. Such new couplings may be the substrate for the long-term storage of information in the brain.

For about 100 years, neuroscientists have been guided by two proposals. The first, by Tanzi², states that the nervous substrate for learning and memory could be a "thickening of the nerve fibres". But Ramón y Cajal³ postulated that this substrate is a strengthening of the connections — or even a building of new connections — between the nerve cells that are most actively engaged in the learning task. Although there is experimental support for these ideas, conclusive evidence has been extraordinarily difficult to find.

A popular cellular model to explain how nervous transmission may be enhanced is long-term potentiation (LTP). This phenomenon is thought to occur either after the synapses to a cell are activated by a train of high-frequency impulses⁴, or by a situation in which the cell receives two simultaneous influences (activation of excitatory

synapses, and an artificially reduced membrane potential, or depolarization), a pairing that is repeated a number of times⁵. Engert and Bonhoeffer¹ have now observed that new sites for communication between neurons — dendritic spines — emerge in a highly restricted part of the dendritic tree of hippocampal cells after LTP in the same region. Because these dendritic spines are the only known postsynaptic partners for the presynaptic boutons in excitatory hippocampal pathways⁶, the findings indicate that new excitatory synapses are formed.

Engert and Bonhoeffer used an organotypic preparation, in which a thin slice of the hippocampus is cultured for several days, and many of the neurons and their connections are preserved. These cells can then be impaled and fluorescently labelled. A group of excitatory synapses on a small dendrite of each neuron was allowed to function inside a tiny sphere, 30 μm in diameter, by superfusion with a fluid similar to that normally found in the brain. The release of neurotransmitters from all other synapses, on this and surrounding cells, was blocked by a solution containing high levels of cadmium and a low concentration of calcium. Local synaptic activation was then paired with depolarizing pulses through an intracellular electrode,



100 YEARS AGO

The old view that insects, with all the lower animals, were created for man's benefit cannot reasonably be held at the present time, but it must, nevertheless, not be forgotten that there are very many beneficial as well as injurious insects. Dr L. O. Howard has recently summed up the good and bad qualities of insects so far as it is possible to do, and he finds that the insects of 116 families are beneficial, and the insects of 113 families are injurious, while those of 71 families are both beneficial and harmful or their functions have not been determined. The injurious insects are made up of 112 families which feed upon cultivated or useful plants, and one family the members of which are parasitic upon warm-blooded animals. Of the beneficial insects, those of 79 families are valuable as preying upon other insects, 32 families are of service as scavengers, two families as pollenisers, and three families as forming food for food fishes.

From *Nature* 4 May 1899.

50 YEARS AGO

In his monograph "Scientific Management" ... Mr. G. Chelioti urges that the essence of management is the art of getting things done through the agency of other human beings, and that management itself is incapable of becoming a science. He regards science as the foundation and the provider of the tools of industry, and the technician as the primary servant of science in industry; but he points out that the technician cannot manage human beings by means of his technology and that the manager dealing with a technical problem becomes a technologist for the time being. In urging this clear separation of the two functions of dealing with human beings and with technical problems or machines, Mr. Chelioti maintains that the undue domination of industry in the nineteenth century by the new element of modern technology, and failure to regard industry as the servant of humanity and to consider sufficiently the human beings employed, was responsible for the revolt of the human spirit against subordination to the machine which we are still experiencing in spite of a more enlightened managerial outlook.

From *Nature* 7 May 1949.

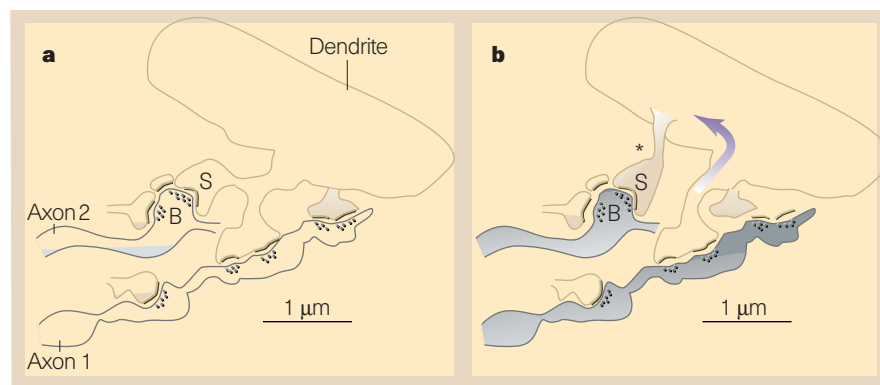


Figure 1 Generation of new dendritic spines in response to long-term potentiation. a, A double-headed spine (S) could theoretically arise by a simple split in an existing spine. In all reconstructed cases, however, such double-headed spines are contacted by two different axons with boutons (B). b, Suggested scheme for the emergence of a new spine (asterisk), according to the results of Engert and Bonhoeffer⁴. The new spine grows towards, and eventually makes contact with, one of the existing boutons of axon 2, an interaction that is triggered, in part, by a signal from the activated spine (curved arrow), connected to axon 1.

and, as expected, LTP was seen in most experiments. Using two-photon microscopy of the impaled and marked cells, the authors observed that new spines emerged about 30 minutes after the LTP was induced. But addition of AP5 (D(-)-2-amino-5-phosphonovaleate), which blocks NMDA (N-methyl-D-aspartate) receptors, prevented the appearance of new spines, indicating that calcium entry is important for this process.

The results are similar to those reported by Maletic-Savatic *et al.*⁷, although these authors used younger organotypic cultures, and a high-frequency stimulation mode which, although not monitored, should elicit LTP in the same neurons. In this case, the changes appeared as filopodia (thin, thread-like protrusions), which are longer and more irregular, but may turn into dendritic spines. Again, appearance of the new projections was highly localized owing to the restricted stimulation, and the filopodia appeared after 20 minutes then continued to grow for another 40 minutes. But, as seen by Engert and Bonhoeffer, they did not appear when AP5 was added, and weak stimulation had no effect on them.

In both reports, the strength of the evidence comes from the highly localized nature of the structural changes — down to a handful of synapses on a small stretch of a dendrite. The thousands of surrounding, naive synapses are left unchanged. Synaptic changes have previously been seen in large networks of cells and synapses, although it is not known how many neurons are involved in these changes. And in the mollusc *Aplysia*, dramatic increases in the number of synapses have been reported after these creatures have learned a conditioned-reflex task⁸. Other examples where new spines have been induced in nerve cells thought to be specifically involved in learning include the lasting avoidance that chicks show after pecking once at a bead doped with a foul taste⁹, spa-

tial training in rats¹⁰, and the response seen after the induction of LTP in anaesthetized rats¹¹.

Exciting (quite literally) as these new findings are, several questions remain. The most pressing is whether the new spines can explain the key element of LTP — the increased postsynaptic current. Are these new spines innervated, and, if so, by which fibres? A convincing answer requires the demonstration, by electron microscopy, that the presynaptic boutons of the stimulated nerve fibres are attached to the new spines, as well as proof that these spines are functionally competent. This could be done by looking at the levels of calcium in the new spine heads in response to synaptic activation.

Another question is this. How similar is the spine growth seen by Engert and Bonhoeffer to relevant data from whole animals? In organotypic cultures, full physiological maturation of synapses is either delayed or absent while considerable synaptic reorganization takes place¹². The observed spine density is, for example, only one-fifth of that seen in the same cells in normal rats¹¹. We cannot be sure about the function of the new filopodia until their relationship to normal dendritic spines is determined. New filopodia have also been seen after nerve fibres synapsing onto cells have been cut, so they could be a response to an unspecific signal associated with high-frequency synaptic activation.

Finally, the mechanisms that underlie the emergence of new spines raise another set of questions. Which signals trigger and guide the new spines? How important is synaptic activation, given that miniature synaptic potentials are necessary to maintain the spines¹³? Does the activity-related increase in the concentration of calcium within the spines¹⁴ — which is important for rapid changes in the shape of the spines¹⁵

and, probably, for growth of existing spines¹⁶ — also regulate the growth of new spines?

We may get some clues from the organization of the two-headed spines that appear after LTP¹¹. In anaesthetized animals, there are far more two-headed spines 30 minutes after induction of LTP than before. In three-dimensional reconstructions, the two heads on one neck never contact boutons of the same axon, contradicting the idea that one spine has simply split into two (Fig. 1a). Instead, the two heads always contact boutons that already serve at least one spine, suggesting that a new spine has grown towards an existing presynaptic bouton (Fig. 1b). If a similar arrangement occurs for Engert and Bonhoeffer's new spines¹, the effect could be seen as a combination of an attractive signal from a neighbouring bouton, coupled with a growth signal from a dendrite next to the activated spine. Whatever the answers, an anatomically precise link between new spines and LTP has been long awaited; a role for LTP and its associated structural changes in behavioural learning may be an even greater challenge. □

Per Andersen is at the Institute for Basic Medical Research, University of Oslo, PO Box 1104 Blindern, N-0317 Oslo, Norway.

e-mail: andersen@basalmed.uio.no

- Engert, F. & Bonhoeffer, T. *Nature* **399**, 66–70 (1999).
- Tanzi, E. *Riv. Sper. Freniat.* **19**, 149 (1893).
- Ramón y Cajal, S. *Proc. R. Soc. Lond. B* **55**, 243–245 (1894).
- Bliss, T. V. P. & Lomo, T. *J. Physiol. (Lond.)* **232**, 331–356 (1973).
- Gustafsson, B., Wigström, H., Abraham, W. C. & Huang, Y. Y. *J. Neurosci.* **7**, 774–780 (1987).
- Andersen, P., Blackstad, T. W. & Lomo, T. *Exp. Brain Res.* **1**, 236–248 (1966).
- Maletic-Savatic, M., Malinow, R. & Svoboda, K. *Science* **283**, 1923–1927 (1999).
- Bailey, C. H. & Chen, M. *Proc. Natl Acad. Sci. USA* **85**, 2372–2377 (1988).
- Patel, S. N., Rose, S. P. R. & Stewart, M. G. *Brain Res.* **463**, 168–173 (1988).
- Moser, M.-B., Trommald, M. & Andersen, P. *Proc. Natl Acad. Sci. USA* **91**, 12673–12675 (1994).
- Trommald, M., Hulleberg, G. & Andersen, P. *Learning and Memory* **3**, 218–222 (1996).
- Protscher, M. & Gähwiler, B. H. *Neuroscience* **24**, 541–551 (1988).
- McKinney, R. A., Capogna, M., Durr, R., Gähwiler, B. H. & Thompson, S. M. *Nature Neurosci.* **2**, 44–49 (1999).
- Korkortian, E. & Segal, M. *Eur. J. Neurosci.* **10**, 2076–2084 (1998).
- Fischer, M., Kaech, S., Knutti, D. & Matus, A. *Neuron* **20**, 847–854 (1998).
- Hosokawa, T., Rusakov, D. A., Bliss, T. V. P. & Fine, A. *J. Neurosci.* **15**, 5560–5573 (1995).

Condensed-matter physics

The magnetic turnabout

Olivier Kahn

The ability of certain pieces of rock to attract or repulse each other was discovered about 25 centuries ago by a Greek shepherd whose name was Magnes. Since that time, magnetic properties of materials have fascinated humankind, and quite a few types of magnets have been characterized, synthesized or manufactured. In *Physical Review Letters*, Hashimoto and co-workers¹ report a unique magnet that reverses magnetization not just once, but twice, as the temperature changes. A simple theoretical approach led to the design and synthesis of this unusual magnet, which is a Prussian blue material incorporating four

different metal ions.

A magnet may be defined as a material exhibiting a spontaneous magnetization (that is, it remains magnetized even when there is zero external magnetic field) below a certain critical temperature, T_c . Some metals, metallic alloys and even molecule-based compounds such as the metallic Prussian blue analogues, behave in this way. The two main kinds of magnets are ferromagnets and ferrimagnets. In ferromagnets, such as iron, interactions between local magnetic dipoles favour a state in which these dipoles align along the same direction, and long-range magnetic ordering occurs at T_c . The sponta-

neous magnetization is zero above T_c , where thermal motion overwhelms any magnetic alignment, then progressively increases as the temperature is lowered down to absolute zero. The spontaneous magnetization at 0 K corresponds to the perfect alignment of the local magnetic dipoles.

In a ferrimagnet, such as magnetite, there are two non-compensating sub-lattices of local dipoles called M_A and M_B , with $M_A \neq M_B$. The interactions between M_A and M_B favour their antiparallel orientation. Again, a long-range ordering occurs at T_c , with the dipoles M_A aligned in one direction and the dipoles M_B oriented in the opposite direction. At any temperature below T_c , the spontaneous magnetization is the algebraic sum of the contributions arising from each of the two sub-lattices. In most cases, at any particular temperature, the contribution of one of the sub-lattices is larger in value, so that the resulting spontaneous magnetization increases steadily as T is lowered (Fig. 1a). In a few cases, the two contributions cancel out at a temperature called the compensation temperature, T_{comp} (Fig. 1b). Between T_c and T_{comp} , the contribution of one of the sub-lattices dominates. Below T_{comp} , the contribution of the other sub-lattice dominates. The spontaneous magnetization is first positive, then zero, and finally negative as the sample is cooled down. Such behaviour was predicted by Louis Néel² in 1948 and later observed experimentally — first in certain metal oxides³, and then in some molecule-based magnets containing organic cations⁴.

If a material contains more than two sub-lattices, a complicated temperature dependence of the spontaneous magnetization may be anticipated, as emphasized in Fig. 1c where two compensation temperatures appear. Hashimoto and co-workers¹ have designed and created such a material using Prussian-blue-like phases. The general formula of these metal-cyanide phases is $A_k[B(CN)_6]_l$, where A and B are transition metal ions with different electron spins. The basic structure is face-centred cubic with A and B ions linked in an alternating fashion. The A ions (Ni, Mn or Fe) are randomly distributed at the A-lattice sites, but are always linked to the nitrogen end of the cyano group (CN), whereas the B-lattice ion (Cr) is always linked to C, forming A–N–C–B linkages in all three perpendicular directions (Fig. 2, overleaf).

With these phases it is possible to use molecular field theory to predict the temperature dependence of the spontaneous magnetization for each of the individual sub-lattices. Such a prediction usually requires complex calculations of interactions between each atom and many of its neighbours, but Hashimoto and co-workers only considered exchanges between nearest neighbours (that is, A–B couplings) because of the relatively long distance between the

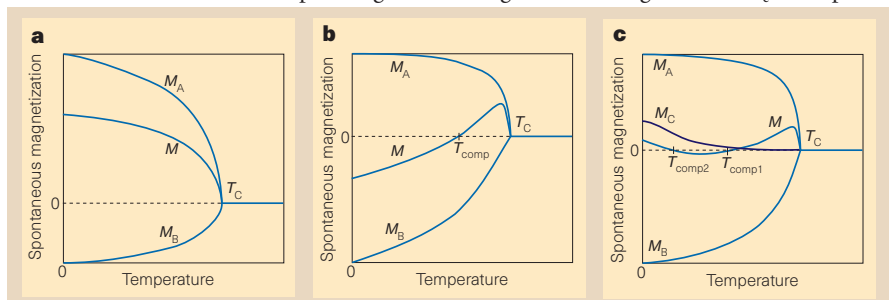


Figure 1 Temperature dependence of the spontaneous magnetization for a ferrimagnet with two sub-lattices. The observed magnetization, M , is the algebraic sum of the contributions arising from each of the sub-lattices, M_A and M_B below a critical temperature T_c . a, Usually M varies monotonically as a function of temperature. b, M shows a compensation temperature, T_{comp} , at which the sign of the magnetization is reversed. c, Temperature dependence of the spontaneous magnetization for a hypothetical magnet with three sub-lattices and two compensation temperatures. The observed magnetization, M , is the algebraic sum of three contributions, M_A , M_B and M_C .

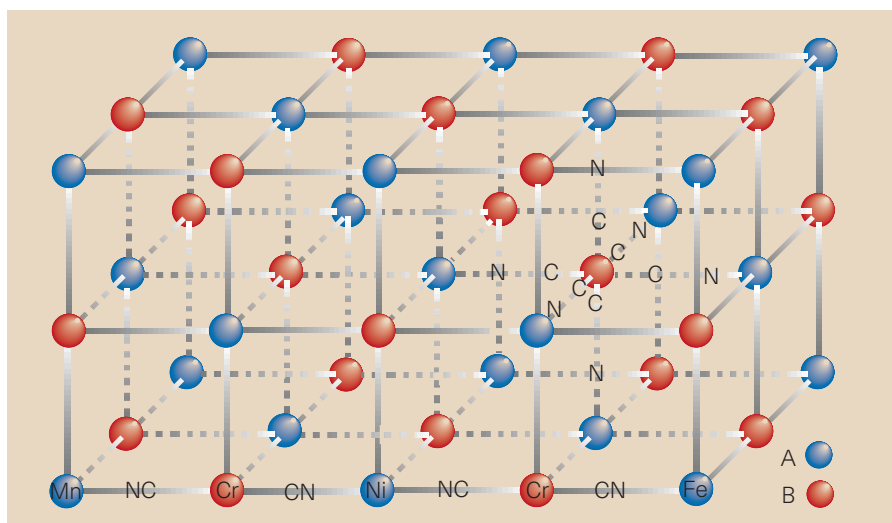


Figure 2 Idealized structure of the Prussian-blue-like phases designed by Hashimoto and colleagues¹. The phase incorporates four transition metal ions (Mn, Cr, Ni and Fe), which produce a mixed ferroferrimagnet capable of switching magnetization twice with changing temperature.

atoms. Moreover, owing to the regularity of the A and B sites and of the lattice as a whole, the symmetry rules governing the magnetic properties apply in a straightforward manner^{5–7}. So the authors were able to predict whether the compound is ferro- or antiferromagnetic, and to estimate T_c .

Hashimoto and co-workers focused on compounds of formula $A_3[Cr(CN)_6]_2$. For A = Ni or Fe, the compound is a ferromagnet, and for A = Mn it is a ferrimagnet. First, the authors synthesized compounds incorporating just three transition ions⁸, with the formula $(Ni_aMn_b)_3[Cr(CN)_6]_2$, where $a + b = 1$. With such materials, they succeeded in tuning the temperature dependence of the spontaneous magnetization by changing the chemical composition as defined by a and b, but not in obtaining two compensation temperatures. In their latest paper¹, they worked with compounds incorporating four transition ions, with the formula $(Ni_aMn_bFe_c)_3[Cr(CN)_6]_2$, where $a + b + c = 1$. The spontaneous magnetizations of the Ni, Fe and Cr sub-lattices are all positive, whereas that of the Mn sub-lattice is negative, producing a mixed ferroferrimagnet.

Using their simplified molecular-field theory calculation, the authors determined a hypothetical composition — that is, a, b and c values — for which the resulting spontaneous magnetization shows two compensation temperatures. Adding an additional positive contribution (Fe ions) results in a positive spontaneous magnetization at temperatures close to absolute zero, and hence a second compensation temperature. Next, they synthesized the material corresponding to this composition, namely $(Ni_{0.22}Mn_{0.60}Fe_{0.18})_3[Cr(CN)_6]_2$. This material is actually a magnet with $T_c = 63$ K, exhibiting two magnetization reversals, at 53 and 35 K. No other compound of this kind has been obtained before, yet the synthesis of this

unique magnet is surprisingly straightforward. It consists of mixing together at room temperature two aqueous solutions, one containing a mixture of Ni, Mn and Fe salts in the ratio a:b:c, the other containing the $[Cr(CN)_6]^{3-}$ complex. The expected material then forms as a powder. More generally, Prussian-blue-like phases are prepared from molecular precursors in very mild conditions, which provides a remarkable flexibility in terms of chemical compensation, and consequently of physical behaviours. Hashimoto and colleagues have also created a Prussian-blue-like phase that reversibly

switches from a paramagnetic to a ferromagnetic state in response to light⁹.

The first two molecule-based magnets were reported in 1986^{10,11}. Since that pioneering work, quite a few research groups have initiated some activity along this line (see for instance ref. 12), and today several dozens of magnets synthesized from molecular precursors are known. The chemists and physicists who work in this field of research are sometimes asked to specify what are the new features arising from the molecular nature of the materials. The rational design of a magnet with two reversals of the magnetization may now be considered as one of the answers. □

Olivier Kahn is in the Laboratoire des Sciences Moléculaires, Institut de Chimie de la Matière Condensée de Bordeaux (ICMCB), 33608 Pessac, France.

e-mail: olivier.kahn@icmcb.u-bordeaux.fr

- Ohkoshi, S., Abe, Y., Fujishima, A. & Hashimoto, K. *Phys. Rev. Lett.* **62**, 1285–1288 (1999).
- Néel, L. *Ann. Phys. (Leipzig)* **3**, 137 (1948).
- Blasse, G. & Gorter, E. W. *J. Phys. Soc. (Jpn. Suppl. B-1)* **17**, 176–182 (1962).
- Mathonière, C., Nuttal, C. J., Carling, S. G. & Day, P. *Inorg. Chem.* **35**, 1201–1206 (1996).
- Kahn, O. *Molecular Magnetism* (VCH, New York, 1993).
- Mallah, T., Thiebaut, S., Verdager, M. & Veillet, P. *Science* **262**, 1554–1557 (1993).
- Kahn, O. *Nature* **378**, 667–668 (1995).
- Ohkoshi, S., Iyoda, T., Fujishima, A. & Hashimoto, K. *Phys. Rev. B* **56**, 11642–11652 (1997).
- Sato, O. *et al. Science* **272**, 704–705 (1996).
- Miller, J. S. *et al. J. Chem. Soc. Chem. Commun.* 1026–1028 (1986).
- Pei, Y., Verdager, M., Kahn, O., Sletten, J. & Renaud, J. P. *J. Am. Chem. Soc.* **108**, 428–429 (1986).
- Proc. 5th Int. Conf. Molecule-Based Magnets (eds Itoh, K., Millar, J. S. & Takui, T.) *Mol. Cryst. Liq. Cryst.* **305**, 306 (1996).

Microtechnology

Laying it on thick

Using an electrochemical fabrication process (called EFAB), Adam Cohen and a team of engineers at the University of Southern California have made several multi-layered structures, such as the 12-layer nickel chain shown here that has six freely moving links. The chain is not much thicker than a sheet of paper, probably making it the world's thinnest. But for the engineers who created it, the most exciting thing about the chain is not its size, but the process used to build it.

The EFAB technique can produce arbitrary 3-D micro-devices in a simple, automated process. EFAB is inspired by the commercial process of 'rapid prototyping', which uses a stacking procedure to build a solid object from layers hundreds of micrometres thick. Devices made by rapid prototyping are usually plastic, whereas EFAB can create metal objects from layers that are 5–8 μm or less.

Each layer is made by depositing a 'sacrificial' material using a high-speed 'instant masking' technique, followed by a



blanket deposit of metal. The layer is levelled, and the process repeats, following a computer-generated pattern. Chemically etching away the sacrificial material leaves behind the desired 3-D structure. Instant masking uses a prefabricated plate that deposits material much like a printing press, making the whole process quicker.

Traditional micro-fabrication requires customized processes, and expensive cleanrooms. Through mass production, EFAB promises to substantially lower the costs of producing many micro-scale devices.

Sarah Tomlin

Palaeoanthropology

Stone legacy of skilled hands

James Steele

In 1964, Leakey, Tobias and Napier¹ added a newly discovered, smaller-brained, stone-tool-using hominid, now dated to 1.8-million years old (Myr), to the genus *Homo* (Fig. 1). They named the new species *Homo habilis*: 'handy man'. The stone tools of that earliest industry, the Oldowan (named after the Olduvai Gorge in Tanzania where *H. habilis* was found), were classified into distinct types that implied their makers had sophisticated technological know-how².

Subsequent work complicated this picture. Experiments in tool manufacture and use showed that the Oldowan tool kit was less organized than the original scheme implied³. There are now at least two other hominid species contending for the title of producers and users of the Oldowan tools, and the case has been made that none of these contenders currently assigned to the genus *Homo* is enough like us to merit the name⁴. However, new findings of very early stone tools west of Lake Turkana in Kenya, described by Roche

*et al.*⁵ on page 57 of this issue, show that their makers — whoever they were, and whatever name we give them — had surprisingly good control over their raw materials.

The archaeological site of Lokalelei 2C, close to the western margin of Lake Turkana, was excavated in 1997. It is dated to 2.34 ± 0.05 Myr, and consists of a dense concentration (about 10 m² in extent) of 2,067 fragments of worked stone, with smaller quantities of rather degraded animal remains. Another 516 artefact fragments were recovered by surface collection. Lokalelei 2C provides a high-resolution record of the activities of hominids who took stone cobbles, mostly of lava, and knapped them for sharp-edged flakes. By fitting the discarded pieces back together — up to 20 from a single cobble — Roche *et al.* show that these hominids had a high degree of control over the force, amplitude and precision of the hand movements required to detach flakes successfully and repeatedly from the parent core (Fig. 2).

The knappers may also have appreciated the mechanical properties of their finer grained raw materials. No one has previously identified such skill so clearly in artefacts that are so old.

In 1991, the first late-Pliocene archaeological site at Lokalelei (Lokalelei 1), also dated to 2.34 ± 0.05 Myr, was excavated as part of the same programme⁶. Analysis of the stone tools found at Lokalelei 1 suggested they had been made by relatively unskilled hands. A plausible inference at the time was that these tools had been made at the very dawn of a tradition of flaked stone tool making⁶, perhaps predating *H. habilis*⁷.

Subsequent findings cast doubt on this view. By 1994, about 1,000 kilometres to the northeast in the Gona River drainage area in Ethiopia, stone tools had been excavated that could be dated to 2.5–2.6 Myr and that had clearly been made by hominids who had mastered the basic knapping skills⁸. Often several overlapping flakes had been removed in succession from the cores. Many flakes were well struck, showing that the makers understood the fracture mechanics of their materials. In 1994, in the adjacent Hadar drainage system, a jawbone similar to that of *H. habilis* associated with typical Oldowan stone flakes was recovered from a layer dated to 2.33 ± 0.07 Myr⁹.

Roche and colleagues' findings⁵ from Lokalelei 2C unequivocally show a good command of basic fracture mechanics, as applied to the raw materials used at this location. Let us not underestimate the difficulty of learning to execute rapid, precise, aimed movements of the arm and hand such as those needed for successful stone flaking. Wild chimpanzees in West African groups with cultural traditions of nut cracking using unmodified stone hammers take several years to become fully proficient at opening the nut without crushing its kernel. The effort is worthwhile: in the nut season, an adult female can obtain 3,800 calories per day this way¹⁰. Understanding, and exploiting, the fracture mechanics of stone itself to produce flaked stone tools adds a new level of complexity to such tasks.

So, which late-Pliocene hominid species could have made these early artefacts, and why? The answer must be found in the fossil traces of their hands and brains. Reconstructing what the tools from these very early sites were used for is frustratingly difficult. Roche *et al.*⁵ found animal remains at Lokalelei 2C, including teeth and bones of grazing mammals, reptiles and fish. However, these remains are poorly preserved, show no tool marks, and may have accumulated at this stream-side location without hominid involvement. Tortoise bones and ostrich egg-shell fragments, found close to stone tools at Lokalelei 2C and 1, may be the remains of hominid meals, but this is an argument from association. We need infor-

Evolution

Heard but not seen

"Speak softly and carry a big stick", said Theodore Roosevelt. But a Darwinian might regard this as bad advice, as demonstrated by W. T. Fitch's explanation of the unusually long tracheae of some birds (*J. Zool. Lond.* 248, 31–49; 1999).

In at least 60 bird species — including the trumpeter swan shown here — the trachea is thrown into coils or loops instead of taking a direct route between the throat and the lungs. The trait shows a puzzling diversity: it is found in six avian orders, and has probably evolved several times. It occurs in migratory cranes and swans, and in large, sedentary rainforest dwellers such as curassows. Closely related species show large variation in trachea length. In some species, only the males possess elongated tracheae, whereas in others both sexes do.

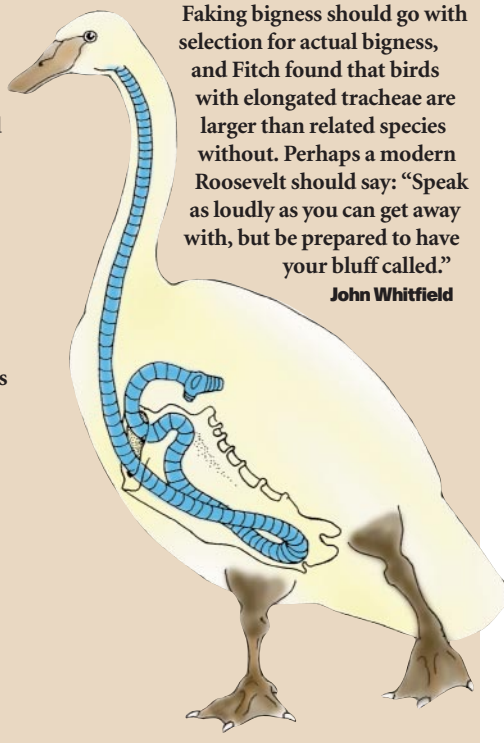
Fitch argues that the best explanation is that a long trachea allows a calling bird to give an exaggerated impression of its size. This works through an acoustic process known as 'reduced formant dispersion': a call produced by an elongated trachea has more closely spaced resonant frequencies, producing a deeper, more baritone sound.

A common feature of birds with tracheal elongation is that they nest in dense vegetation, where visibility is poor, and where duplicating the call of a larger bird will be useful in defending a

territory, or possibly attracting a mate. The lack of visual cues reduces the possibility that their trick will be discovered, but one would expect natural selection to promote scepticism in listeners. This might not happen, however, if honest calls remained in a sufficiently large majority.

Faking bigness should go with selection for actual bigness, and Fitch found that birds with elongated tracheae are larger than related species without. Perhaps a modern Roosevelt should say: "Speak as loudly as you can get away with, but be prepared to have your bluff called."

John Whitfield



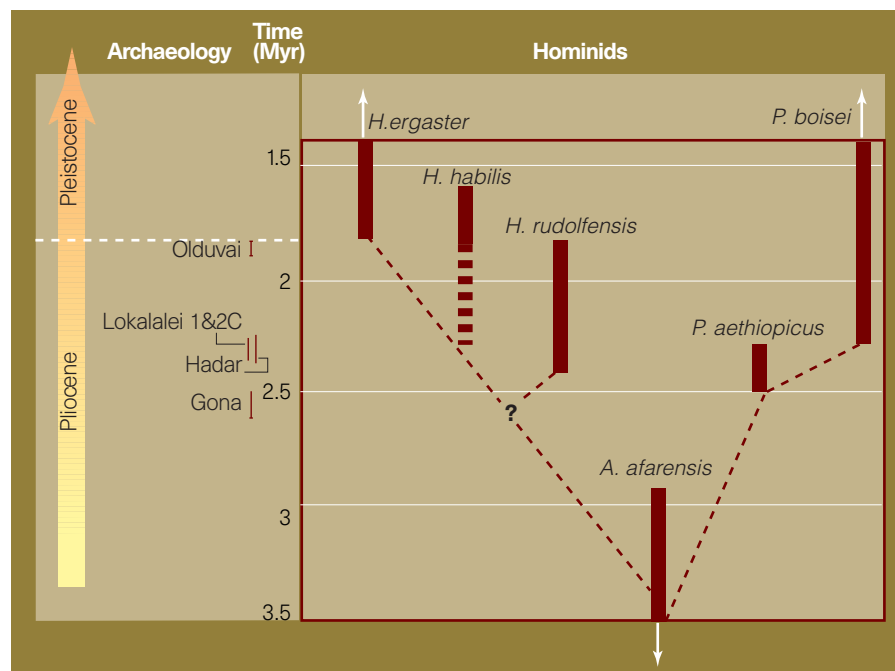


Figure 1 Absolute ages of the archaeological sites described in the text, and of the hominid species extant in East Africa 3.5–1.5 million years ago (Myr). Dashed lines show probable evolutionary relationships. Olduvai, site FLKNN1; Hadar, site AL666. (Modified from ref. 16). New fossils from Ethiopia¹⁷ may provide the missing link at the roots of the genus *Homo*, and date from 2.5 Myr.



Figure 2 'Refit' from Lokalalei 2C. Roche *et al.*⁵ reconstructed the arrangement of sharp-edged flakes, used as tools, that were struck from a core stone. Scale bar is 10 cm long.

mation from other sites, predating the two-million year mark, that have both stone tools and tool-marked bones, before reaching any firmer conclusions.

Meanwhile, no one is certain who wielded these tools. Using electromyography, Marzke and colleagues¹¹ identified the hand muscles crucial in the experimental manufacture of Oldowan tools. Habitual tool making would be reflected in the bones of the hand regions stressed by these muscles, and in joints whose configurations affect these muscles' biomechanical efficiency. The findings of Roche *et al.*⁵ will prompt a renewed examination of fossil evidence for the evolution of hominid hands.

Manual dexterity in primates is also correlated with specific aspects of brain organization^{12,13}. So far, the best quantitative measures of early-hominid brain organization relate to cranial capacity. This is a crude measure, but not useless¹³. The average cra-

nial capacities of adult great apes range from 393 to 465 cm³ (ref. 14). Hominid species living 2.5–2.3 Myr ago included early robust australopithecines (*Paranthropus aethiopicus*, cranial capacity about 410 cm³), and the earliest *Homo* (with affinities to *H. habilis*, 500–650 cm³, and/or *H. rudolfensis*, 600–800 cm³; ref. 15). Thirty-five years on, can we emulate Leakey *et al.*¹, and place the tools at Lokalalei into the hand of *Homo*, using an argument from design?

The new findings⁵ will not resolve this issue. But they bring such a resolution closer with hard evidence of the considerable technical skill of at least some late-Pliocene hominids.

James Steele is in the Department of Archaeology, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

e-mail: tjms@soton.ac.uk

1. Leakey, L., Tobias, P. & Napier, J. *Nature* **202**, 7–9 (1964).
2. Leakey, M. D. *Olduvai Gorge, Vol. 3: Excavations in Beds I and II 1960–1963* (Cambridge Univ. Press, 1971).
3. Toth, N. *J. Archaeol. Sci.* **12**, 101–120 (1985).
4. Wood, B. & Collard, M. *Science* **284**, 65–71 (1999).
5. Roche, H. *et al.* *Nature* **399**, 57–60 (1999).
6. Kibunjia, M. *J. Hum. Evol.* **27**, 159–171 (1994).
7. Kibunjia, M., Roche, H., Brown, F. H. & Leakey, R. E. *J. Hum. Evol.* **23**, 432–438 (1992).
8. Semaw, S. *et al.* *Nature* **385**, 333–336 (1997).
9. Kimbel, W. H. *et al.* *J. Hum. Evol.* **31**, 549–561 (1996).
10. Boesch, C. in *Tools, Language and Cognition in Human Evolution* (eds Gibson, K. R. & Ingold, T.) 171–183 (Cambridge Univ. Press, 1993).
11. Marzke, M. *et al.* *Am. J. Phys. Anthropol.* **105**, 315–332 (1998).
12. Lemon, R. N. *Neuroscientist* **3**, 389–398 (1997).
13. Steele, J. *Curr. Psychol. Cognition* **17**, 1202–1214 (1998).
14. Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Chapman & Hall, London, 1989).
15. Conroy, G. C. *Reconstructing Human Origins* (Norton, New York, 1997).
16. Wood, B. *BioEssays* **18**, 945–954 (1996).
17. Aswaf, B. *et al.* *Science* **284**, 629–635 (1999).

Daedalus

Adhesive antibodies

Many alternative medicines are said to 'boost the immune system'. Few orthodox drugs can make this claim. (One possible exception, the veterinary anti-helminthic Ivermectin, has discouraging side-effects). Daedalus has been pondering the matter.

Our immune system responds to an invading antigen by raising antibodies, which bind selectively to the antigen molecules. The bound molecules usually couple together in their turn to give a big, stable immunocomplex particle, easily recognized and swallowed by the body's scavenging macrophage cells. By contrast, the antibodies of a poor immune system may bind only loosely to the antigen, or fail to couple further. The resulting small or fragile agglomerations are far harder to mop up, and may cause trouble of their own. So Daedalus aims to tighten the binding, with some sort of selective molecular glue.

Chemists will immediately recognize this concept. Many substances crystallize only with 'water of crystallization'. The water molecules fill awkward gaps in the crystal lattice, giving a tightly packed and stable solid. Other small molecules can also work. With the aid of the right glue molecule, even the feeblest antibody might bind firmly to an invading antigen.

DREADCO's biochemists are finding this idea rather a challenge. Antibody–antigen binding is almost impossible to predict; furthermore, different species, or even different strains of the same species, can raise quite different antibodies to a given antigen. Indeed, says Daedalus, many quirky nostrums or remedies may occasionally work only because they contain some small glue-molecule that happens to fit the patient's individual antibodies. So his team is devising a 'cocktail' of small, polar molecules, to be injected into the body like a charge of shotgun pellets. With luck, one or more of them will act as glue molecules for the current antibodies.

Urea, DMSO, alcohol and acetone are obvious candidates, as are the more polar fluorohydrocarbons. Heavy water can influence protein folding; it, and deuterio-versions of the other molecules, will also be included. In particular, Daedalus has high hopes of deuteriohydrogen, HD. Its small size will let it sneak into tiny gaps in the binding, and its asymmetry will give it a bit of useful polarity. A deuteriohydrogen breathing chamber could do wonders for an overburdened immune system. Sadly, the mixture with oxygen will be devastatingly explosive.

David Jones

Obituary

David Phillips (1924–99)

A founding father of structural biology

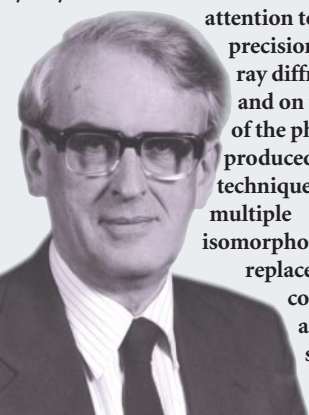
Nineteen sixty was the spring of hope for protein crystallography. The determination of the structures of myoglobin at 2-Å resolution by John Kendrew and of haemoglobin at 5.5 Å by Max Perutz had shown how structural studies could yield biological information. Yet, for several years, no new protein structures were solved and anxiety set in. Maybe the globins had been flukes.

David Phillips's structure determination of lysozyme in 1965 completely changed that view. This work first showed the power of protein crystallography to explain biological function in terms of physics and chemistry. It changed forever the way that enzymes are studied and was the start of the explosion in the number of protein structures being solved.

Following a war-time undergraduate degree in physics, mathematics and electrical engineering at Cardiff University, and subsequent appointments in Cardiff and Canada, Phillips took a post-doctoral position at the Royal Institution, London, in 1956. Sir Lawrence Bragg was director and fostered a spirit of independence and initiative there. Realizing that automating the collection of X-ray diffraction data was a prime objective for studies of large protein molecules, one of Phillips's first tasks was to collaborate in the design and construction of an automated diffractometer, an analogue device based on a mechanical model of the diffraction lattice. With this instrument he and his team were able to obtain highly accurate data which, in turn, led to precise structures (including that of myoglobin at 1.4-Å resolution, an astonishing precision for that era).

The solution of the X-ray structure of lysozyme was based on meticulous

attention to the precision of the X-ray diffraction data and on the quality of the phases produced by the technique of multiple isomorphous replacement coupled with anomalous scattering



measurements. The structure showed the complete path of the polypeptide chain (129 amino-acid residues) folded into both α -helices, as seen in myoglobin, and β -sheet, a structure that had been predicted by Linus Pauling but not hitherto observed in three dimensions.

Lysozyme has antibacterial activity, which was then known to be due to its ability to hydrolyse the polysaccharide component of the bacterial cell wall. Even in 1962 Phillips had been thinking about the catalytic mechanism, and suggested that, as a graduate project, one of us (Louise Johnson) should pursue binding studies on inhibitor molecules. By early 1966 these experiments had led to a detailed interpretation of lysozyme in complex with an inhibitor, tri-*N*-acetylchitotriose, and an understanding of the key elements by which lysozyme recognizes sugars. The next step was to work out precisely how lysozyme recognizes its hexasaccharide substrate in the bacterial cell wall. By molecular model building, and bringing all the available biochemical evidence to bear, Phillips produced a proposal for the way in which a substrate must bind — and, after further intense thought and insights from Charles Vernon, for lysozyme's catalytic mechanism.

Here was the first structural explanation of how an enzyme speeds up a chemical reaction. The proposals contained suggestions for essentially every structural contribution to the catalytic power of enzymes that has since been detected — proximity, ground-state distortion, alteration of catalytic-group pK_a by the unique environment of the enzyme, general acid–base catalysis, and transition-state stabilization by electrostatic and hydrogen-bonding interactions. The extrapolation from inhibitor binding to substrate binding was a remarkable feat of deductive reasoning. It was achieved in a mere three days — three days which Phillips described as the most rewarding he had ever spent.

In 1966, he was appointed professor of molecular biophysics at Oxford University, and fresh achievements followed. Most notably, with his graduate students he solved the structure of triose phosphate isomerase (TIM). This was the first example of an eight-fold α - β -barrel protein. David was too modest to call it a 'Phillips fold' but he took quiet pride whenever a new TIM-barrel (as it is now known) structure was announced. The number is now in the hundreds.

From about the mid-1970s Phillips began his second career, in science policy and administration. Scientific research, he believed, must be organized so that "Combined with the provision of the necessary infrastructure, it can release individual scientists to display their critically important gifts of spontaneity and originality". These were his goals when, from 1983 to 1993, he was chairman of the Advisory Board for the Research Councils (ABRC), the then intermediary body between government and the research councils.

His time at ABRC was not without controversy. On the one hand he had to deal with the demands for funding from scientists faced with the growth of scientific opportunities, the need for increasingly complex apparatus and facilities, and the growing importance of interdisciplinary research. On the other hand he fought to persuade government to deliver more money while recognizing the necessarily limited resources and pressures for concentrating them. He won the respect of both sides, emphasizing that only the best science should be funded, although some of his views on priorities were not generally appreciated at the time. The Member of Parliament Tam Dalyell has recounted that politicians were much in awe of him: one minister confided, "I read my brief three times before Phillips enters my office".

David Phillips was appointed a life peer in 1994. He chose to keep his surname for the title ("Otherwise one disappears behind a different name and nobody quite knows who you are"); but, because there had been a Lord Phillips before, he became Lord Phillips of Ellesmere, after his birthplace, a small town in Shropshire close to the Welsh border.

Even when cancer had him deep in its grip, David remained at work and he completed a manuscript, "How the lysozyme molecule was actually solved", just nine days before his death on 23 February of this year. As Conan Doyle has Watson say in *The Final Problem*: "This was one of the best and wisest men we shall ever know". He was a man of remarkable warmth and sensitivity — a true friend in the best sense of the word.

Louise N. Johnson and Gregory A. Petsko

Louise N. Johnson is in the Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3QU, UK.

e-mail: louise@biop.ox.ac.uk

Gregory A. Petsko is at the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, 415 South Street, Waltham, Massachusetts 01254-9110, USA.

e-mail: petsko@binah.cc.brandeis.edu

Naked foraminiferans revealed

It is generally assumed that the first fossil appearance of a group of organisms corresponds to its evolutionary origin. But we have molecular evidence that extant members of the most abundant microfossil-forming group, the Foraminifera, include 'naked' amoeboid species, indicating that ancestral foraminiferans could be unfossilized. This means that the origin of the group might be much earlier than has been deduced from the fossil record. This might help to explain the conflicting molecular and fossil data on the origin of the Foraminifera.

Foraminifera are classically defined as marine, shelled protists. Although the foraminiferan fossil record extends to the Cambrian period¹, molecular data² indicate a much earlier origin for these important microfossils. This disparity suggests that microfossils do not provide adequate evidence for the origin of this group. Such a gap in the fossil record could be explained by the existence of non-skeletonizing ('naked') foraminiferans that were not preserved in sediments. However, as foraminiferans did not include a naked species, we looked for the potential foraminiferan ancestor within the class Athalamia, a sister group to the Foraminifera³.

We have obtained ribosomal DNA and actin gene sequences from *Reticulomyxa filosa*, a well-known "giant freshwater amoeba"⁴ that lacks a shell but possesses pseudopodia that are strikingly similar to those of foraminiferans⁵. Phylogenetic analyses of these molecular data show that *R. filosa* is actually a foraminiferan, indicating that certain foraminiferans might lack a shell and live in freshwater environments.

The complete small subunit (SSU) rDNA sequences were obtained from *R. filosa* and five foraminiferan species. The SSU rDNA of *R. filosa* shows several common features with other foraminiferans, including a very low G + C content (32.6%) and several foraminiferan-specific insertions located in the conserved regions of the gene that make it unusually long (3,347 nucleotides). Phylogenetic analyses of SSU rDNA sequences, using different computational methods, consistently show that *Reticulomyxa* branches within the Foraminifera (Fig. 1a). Its position within the Foraminifera is not well established, but a comparison of more than 100 foraminiferan partial SSU rDNA sequences (not shown) suggests that the genus is most closely related to some primitive thecate Allogromiida.

These results were confirmed by an

analysis of foraminiferan actin-coding genes. Two actin gene families were found in *Reticulomyxa* as well as in the two foraminiferans examined (*Allogromia* sp. and *Ammonia* sp.). Both actin gene sequences branch next to the Euglenozoa in the lower part of the phylogenetic tree (Fig. 1b) in a position remarkably similar to that suggested by analysis of the SSU rDNA data. Each actin family forms a clade supported by a high bootstrap value (98% and 100%) in which *Allogromia* appears as a sister group to *Reticulomyxa* and *Ammonia*. These relationships are supported by high bootstrap values (92% and 96%) and remain stable in all analyses based on amino acid and DNA sequences.

The placement of *Reticulomyxa* within the Foraminifera leads us to challenge the micropalaeontological bias imparted to this group: the Foraminifera can no longer be defined strictly as shelled protists. Indeed, their only taxonomically consistent morphological character is the presence of granuloreticulopodia. Although *R. filosa* probably lost its shell secondarily as an adaptation to the freshwater environment, our findings nevertheless suggest that ancestral foraminiferans could have been naked and existed well before the dawn of skeletonization. This hypothesis helps to reconcile the conflicting molecular and fossil data on the origin of the Foraminifera. Our data also provide support for recent molecular studies^{6,7} that implicate a prolonged period of Precambrian diversification among the major eukaryotic groups.

We predict that naked foraminiferans similar to their hypothetical ancestors might still be living in the marine environment. Finding these species is essential for future molecular studies of the group.

Jan Pawlowski*†, Ignacio Bolivar*, José Fahrni*, Colomán de Vargas*, Samuel S. Bowser‡

*Department of Zoology and Animal Biology, University of Geneva,

1224 Chêne-Bougeries, Geneva, Switzerland
e-mail: pawlow@sc2a.unige.ch

†Museum of Natural History,
1211 Geneva 4, Switzerland

‡Wadsworth Center,
New York State Department of Health,
P.O. Box 509, Albany, New York 12201, USA

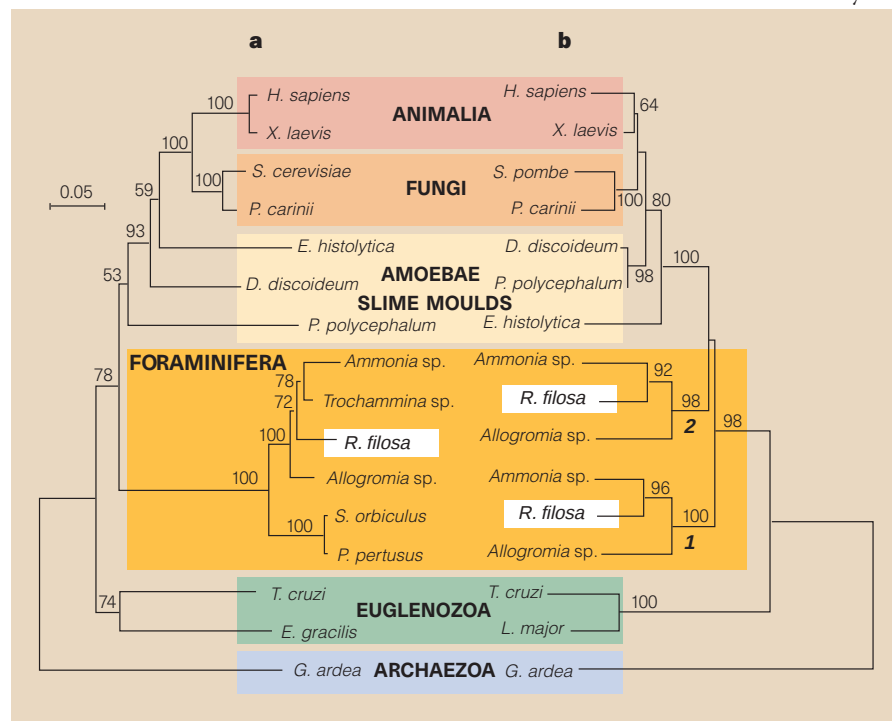


Figure 1 Phylogenetic position of *Reticulomyxa filosa* and foraminiferans. The position was inferred from complete sequences of SSU rDNA (**a**) and actin amino acids (**b**), using the neighbour-joining method with Kimura 2 and Dayhoff's PAM (Percent Accepted Mutations) matrix distances, respectively. The two foraminiferan actin gene families are numbered 1 and 2. The scale bar corresponds to the number of substitutions per site. The GenBank accession numbers of new sequences reported here are AJ 132367 to AJ 132375.

1. Culver, S. J. *Science* **254**, 689–691 (1991).
2. Pawlowski, J. et al. *Mol. Biol. Evol.* **14**, 498–505 (1997).
3. Lee, J. J., Hutner, S. H. & Bovee, E. C. (eds) *An Illustrated Guide to the Protozoa* (Allen, Lawrence, Kansas, 1985).
4. Koonce, M. P., Euteneuer, U. & Schliwa, M. *J. Cell Sci.* (suppl.) **145**–159 (1986).
5. Travis, J. L. & Bowser, S. S. in *Biology of Foraminifera* (eds Lee, J. J. & Anderson, O. R.) 91–156 (Academic, London, 1991).
6. Cooper, A. & Fortey, R. *Trends Ecol. Evol.* **13**, 151–156 (1998).
7. Bromham, L., Rambaut, A., Fortey, R., Cooper, A. & Penny, D. *Proc. Natl Acad. Sci. USA* **95**, 12386–12389 (1998).

A stimulatory phalloid organ in a weaver bird

The males of very few bird species possess a penis¹. Buffalo weavers, *Bubalornis* spp., are unique in possessing a false penis, or phalloid organ^{2,3}. We find that, after protracted copulation, the phalloid organ generates an orgasm-like state in males, a phenomenon not known in any other bird. The possession of a phalloid organ and orgasm may be associated with the buffalo weaver's unusual mating system in which there is intense sperm competition.

We studied individually colour-ringed red-billed buffalo weavers, *B. niger* (Fig. 1a), in Namibia. The phalloid organ is a stiff rod of connective tissue 15.7 ± 0.29 (s.e.) mm long ($n = 109$) that lies immediately anterior to the cloaca (Fig. 1b). It has no ducts^{2,3} and is not homologous to any penis found in any other bird species¹. Females have a much smaller phalloid organ (6.1 ± 0.19 mm long, $n = 68$).



Figure 1 Buffalo weavers, *Bubalornis niger*, have a false penis, or phalloid organ. **a**, Head of a male buffalo weaver. **b**, A male held to show the phalloid organ. **c**, A pair of captive buffalo weavers copulating.

The buffalo weavers bred in colonies of 5.7 ± 1.1 nests ($n = 24$), with each nest containing 6.5 ± 0.8 ($n = 16$) chambers, of which 4.6 ± 0.7 ($n = 13$) were occupied by a single female. Of ten nests studied, two were defended by one male, and eight by two males who appeared to operate as a coalition: both males built the nest, defended it against other males, and fed the chicks in all nest chambers. The mating system was therefore either polygyny or, at two-male nests, cooperative polygyny.

Multilocus DNA fingerprinting⁴ of ten male coalitions revealed that in only one instance were males probably first-order relatives (brothers, or father and son); in at least seven other cases, coalition males appeared to be completely unrelated. Sperm competition was intense. Multilocus DNA fingerprinting of 55 offspring from 25 broods indicated that, in 18 broods (72%), offspring either had more than one father or included offspring fathered by non-resident males.

Copulations were difficult to observe because they occurred between 75 and 500 metres from the nest in dense tree cover. Six complete sequences were protracted (mean 18.0 min, range 4–30 min) and comprised alternating sequences of 'bouncing' displays by both sexes and mounting. It was impossible to establish the position or role of the phalloid organ during copulation as it was obscured by feathers. We therefore caught (under licence) 13 males and 6 females, transported them to Germany and kept them in aviaries, where we observed 57 copulations between three males and five females. In captivity, males also copulated 34 times with a taxidermist's mounted female with an artificial cloaca⁵ (a 'model female'), which we placed inside the aviary.

Mounting was similar to that in other birds, with the male on the female's back. But as copulation proceeded, the male usually adopted a reclining position, leaning backwards away from the female (Fig. 1c). Contrary to earlier speculation^{1,3}, the phalloid organ was not intromittent, but instead appeared to be rubbed against the female's cloacal region. As in the wild, mounting bouts were prolonged (29.0 ± 0.76 min, $n = 34$). In the final stages of copulation, males appeared to experience an orgasm during which the male's wing beat slowed to a quiver and his entire body shook; with his leg muscles apparently in spasm, his feet clenched hold of the female and drew her towards him. The behaviour of males copulating with the model female was similar, and only in those instances when males experienced an orgasm ($n = 34$) was semen found in the false cloaca, suggesting that orgasm is necessary for ejaculation.

To test the idea that protracted mounting and physical stimulation of the phalloid organ results in orgasm and ejaculation, we

rubbed the phalloid organ of each of three males (on two separate occasions each) immediately after they had copulated with and inseminated a live female. All three ejaculated almost immediately each time; similar stimulation without mounting was ineffective.

Our results indicate that the phalloid organ is a stimulatory organ that may have evolved in response to the intense sperm competition mediated, in part, by a polygynandrous mating system. Unusual copulation behaviour occurs in other birds with intense sperm competition^{6,7}. The way in which a stimulatory phalloid organ increases male reproductive success remains unclear, however. One possibility is that, as in some rodents⁸, repeated mounting and protracted physical stimulation of the male, and consequently the female, before ejaculation increases the likelihood that a female will retain and utilize his sperm.

M. Winterbottom, T. Burke, T. R. Birkhead

Department of Animal and Plant Sciences,
University of Sheffield,
Western Bank, Sheffield S10 2TN, UK
e-mail: t.r.birkhead@sheffield.ac.uk

1. King, A. S. in *Form and Function in Birds* (ed. King, A. S. & McLelland, J.) 107–147 (Academic, London, 1981).
2. Bentz, G. D. *Auk* **100**, 501–504 (1983).
3. Birkhead, T. R., Stanback, M. T. & Simmons, R. E. *Ibis* **135**, 326–331 (1993).
4. Burke, T. & Bruford, M. W. *Nature* **327**, 149–152 (1987).
5. Pellatt, E. J. & Birkhead, T. R. *Ibis* **136**, 97–101 (1994).
6. Davies, N. B. *Dummock Behaviour and Social Evolution* (Oxford Univ. Press, 1992).
7. Schulze-Hagen, K., Leisler, B., Birkhead, T. R. & Dyrce, A. *Ibis* **137**, 85–91 (1995).
8. Matthew, M. K. & Adler, N. T. *Physiol. Behav.* **20**, 303–309 (1978).

Green processing using ionic liquids and CO₂

Many organic solvents evaporate into the atmosphere with detrimental effects on the environment and human health. But room-temperature ionic liquids, with low viscosity and no measurable vapour pressure¹, can be used as environmentally benign media for a range of industrially important chemical processes^{2–6}, despite uncertainties about thermal stability and sensitivity to oxygen and water. It is difficult to recover products, however, as extraction with water⁷ works only for hydrophilic products, distillation is not suitable for poorly volatile or thermally labile products, and liquid–liquid extraction using organic solvents results in cross-contamination. We find that non-volatile organic compounds can be extracted from ionic liquids using supercritical carbon dioxide, which is widely used to extract large organic compounds with minimal pollution⁸. Carbon dioxide dissolves in the liquid to facilitate extraction, but the ionic

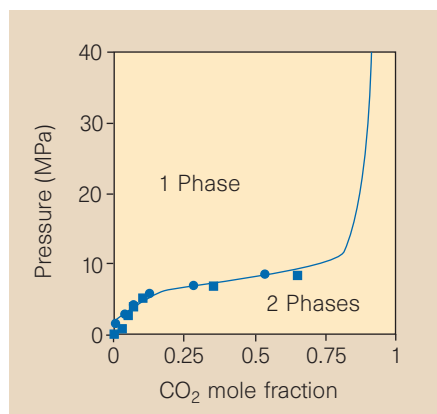


Figure 1 Phase diagram for the CO_2 -[BMIM][PF₆] system. CO_2 solubility in the ionic liquid-rich liquid phase at CO_2 mole fractions up to 0.75 was determined using a high-pressure view cell[†] rated to 8.3 MPa. The number of moles of CO_2 in the liquid phase at a given temperature and pressure was determined by difference, using IUPAC pure CO_2 density data¹² to obtain the number of moles of CO_2 in the vapour phase (assumed to be pure CO_2). Cloud points of dilute mixtures (1.3–7.2% ionic liquid) were determined using a variable-volume view cell[†] rated to 40 MPa. The solubility of the ionic liquid in the CO_2 -rich phase was checked with an ISCO 220SX high-pressure extraction apparatus. A small quantity of ionic liquid was loaded onto a polyurethane foam sponge and extracted with CO_2 at 13.8 MPa and 40 °C. The saturated CO_2 phase leaving the system was depressurized through a heated restrictor and slowly bubbled through a flask of ethanol. Ultraviolet and visual absorption of the collection solvent was used to determine that no ionic liquid was extracted. The circles and squares indicate two separate, replicate experiments.

liquid does not dissolve in carbon dioxide, so pure product can be recovered.

We synthesized⁷ the ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate [BMIM][PF₆], which is stable in the presence of either oxygen or water. Our primary objective was to show that CO_2 could be used to extract naphthalene, our low-volatility model solute, from an ionic liquid, but it was important to show that the CO_2 -rich phase is not significantly contaminated by the ionic liquid, as would be expected during contact of CO_2 with any conventional organic solvent. We therefore investigated the phase behaviour of [BMIM][PF₆] with CO_2 , as well as with naphthalene, and finally that of the [BMIM][PF₆]- CO_2 -naphthalene ternary.

CO_2 is highly soluble in [BMIM][PF₆] (Fig. 1), reaching a mole fraction of 0.6 at 8 MPa, yet the two phases are not completely miscible: cloud points of mixtures ranging from 1.3 to 7.2 mole % of ionic liquid at 25 °C could not be found at pressures up to 40 MPa, the highest pressure accessible with our equipment. The composition of the CO_2 -rich phase is essentially pure CO_2 . After extracting the ionic liquid with 55 g

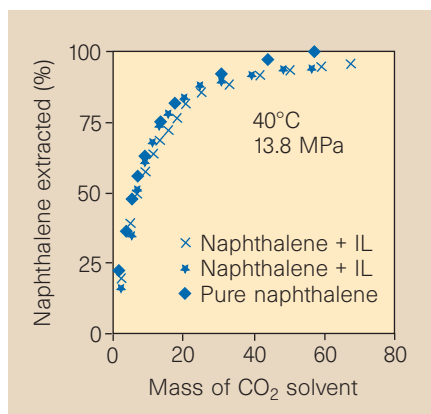


Figure 2 Extraction of naphthalene from the naphthalene/[BMIM][PF₆] mixture using CO_2 at 40 °C and 13.8 MPa. The ISCO extractor was the same as that in Fig. 1. The naphthalene concentration in the ethanol collection solution was determined by ultraviolet and visual spectroscopy, using peaks at 267, 276 and 285 nm. Replicate experiments indicate 96% recovery (crosses) from a 0.0431 g sample (containing 0.00265 g naphthalene) contacted with 67 g CO_2 , and 94% recovery (stars) from a 0.0409 g sample (containing 0.00252 g naphthalene) contacted with 56 g CO_2 . Data for the dissolution of an equivalent amount of solid naphthalene into CO_2 , using the same extractor and analytical technique, are also shown (diamonds).

CO_2 at 13.8 MPa and 40 °C, there was no detectable [BMIM][PF₆] in the extract, indicating that the solubility is less than 10^{-5} mole fraction. In contrast, a mixture of CO_2 with a conventional organic liquid results in significant solubility of the liquid in the CO_2 -rich phase. The phase behaviour of the ionic liquid- CO_2 system resembles that of a cross-linked polymer-solvent system⁹, even though [BMIM][PF₆] is a low-viscosity, low-molecular-weight liquid. The liquid phase increased in volume by only 10–20% when 8 MPa of CO_2 pressure was applied, perhaps because it is ionic; this corresponds to a more than twofold decrease in the molar volume.

Naphthalene was chosen as our model non-volatile organic solute because it dissolves readily in [BMIM][PF₆] (maximum solubility of 0.30 mole fraction at 40 °C) and in CO_2 (with a solubility of 0.013–0.017 mole fraction at 35 °C and pressures of 12.2–20.4 MPa; ref 10). A mixture of 0.12 mole fraction naphthalene in [BMIM][PF₆] was extracted with CO_2 at 13.8 MPa and 40 °C with recoveries of 94–96% (Fig. 2). This near-quantitative recovery compares favourably with the dissolution of a similar amount of pure solid naphthalene with comparable amounts of CO_2 (Fig. 2). It is therefore possible to quantitatively extract a non-volatile organic solute from an ionic liquid using CO_2 without any contamination. Moreover, the dissolution of CO_2 in the ionic liquid is completely reversible: pure ionic liquid

remains after extraction of the naphthalene and depressurization.

Lynnette A. Blanchard*, Dan Hancu†, Eric J. Beckman†, Joan F. Brennecke*

*Department of Chemical Engineering,
University of Notre Dame,
Notre Dame, Indiana 46556, USA
e-mail: jfb@nd.edu

†Department of Chemical Engineering,
University of Pittsburgh,
Pittsburgh, Pennsylvania 15261, USA

1. Seddon, K. R. *J. Chem. Tech. Biotechnol.* **68**, 351–356 (1997).
2. Chauvin, Y. & Olivier-Bourbigou, H. *Chemtech* **25**, 26–30 (1995).
3. Adams, C. J., Earle, M. J., Roberts, G. & Seddon, K. R. *Chem. Commun.* **19**, 2097–2098 (1998).
4. Earle, M. J., McCormac, P. B. & Seddon, K. R. *Chem. Commun.* **20**, 2245–2246 (1998).
5. Earle, M. J., McCormac, P. B. & Seddon, K. R. *Green Chem.* **1**, 23–25 (1999).
6. Ellis, B., Keim, W. & Wasserscheid, P. *J. Chem. Commun.* **4**, 337–338 (1999).
7. Huddleston, J. G. *et al. Chem. Commun.* **16**, 1765–1766 (1998).
8. Brennecke, J. F. *Chem. Ind.* **21**, 831–834 (1996).
9. Moerkkerke, R. *et al. Macromolecules* **31**, 2223–2229 (1998).
10. Sako, S., Ohgaki, K. & Katayama, T. *J. Supercrit. Fluids* **1**, 1–6 (1988).
11. Huie, N. C., Luks, K. D. & Kohn, J. P. *J. Chem. Eng. Data* **18**, 311–313 (1973).
12. Angus, S., Armstrong, B. & de Reuck, K. M. (eds) *International Thermodynamic Tables of the Fluid State: Carbon Dioxide* (Pergamon, Oxford, 1976).
13. Super, M. & Beckman, E. J. *J. Chem. Eng. Data* **42**, 664–670 (1997).

Long-distance transport of pollen into the Arctic

Airborne particulates can be carried over long distances, but for significant quantities of particulates larger than a few micrometres in diameter to be transported more than a few kilometres usually requires a means of injecting the material high into the atmosphere, such as a volcanic eruption, forest fire or desert windstorm. But an unusual event occurred in the Canadian Arctic last year, in which significant amounts of pine and spruce pollen (30–55 µm long) were transported roughly 3,000 km.

On the night of 5 June 1998 and the following morning, local hunters noticed an unusual concentration of pollen at the edges of ponds on Arctic Ocean ice near Repulse Bay, Northwest Territories, Canada (Fig. 1). Similar deposits were also reported in Pelly Bay, northwest of Repulse Bay, but not in other Arctic communities. The material is a remarkably pure pollen concentrate containing 92% jack pine (*Pinus banksiana*) and 8% white spruce (*Picea glauca*). An examination of more than 1,500 pollen grains revealed no pollen from alder (the most common exotic pollen in the eastern Arctic¹) or other taxa, no large organic debris and no charcoal. The pollen is in excellent condition, most of it having unbroken walls and intact cytoplasm, unlike the ‘yellow rain’ reported in south-

east Asia in the late 1970s and early 1980s, which proved to be bee faeces².

Pine pollen is known to traverse great distances, and is often found as a component of both pollen rain and stratigraphic pollen assemblages well beyond the tree line³. Spruce pollen, being much larger, does not transport as readily but can nevertheless form a significant proportion of the tundra pollen assemblage³. Typically, boreal forest pollen deposit in the high Arctic at a rate of less than 1 grain per species per cm² per year¹. The influx rate in Repulse Bay during this event was not measured, but must have been of the order of hundreds of grains per cm² in only a few hours.

The timing of the deposit and its geographic coverage indicate that the event was the result of an unusually strong low-pressure system that developed over Repulse Bay on 5 June. A three-dimensional back-trajectory analysis indicates that winds arriving at Repulse Bay on 6 June would have been near ground level in central Quebec on 1 June. Pine, spruce and alder may all be in flower at this time of year, depending on weather conditions⁴. High surface winds (up to 24 km h⁻¹) lofted the pollen into the air, and this pollen-laden air mass travelled at approximately the 850 hPa level

(~1,300 m) northeast over Labrador, north over the Labrador Sea, and west over southern Baffin Island, arriving at Repulse Bay late on 5 June and in the morning of 6 June. The wind speed dropped to less than 20 km h⁻¹, allowing the pollen to settle out. This event lasted an unusually long time (turbulence in the airflow causes most grains to remain airborne for less than a day⁵) and covered a remarkable distance (nearly 3,000 km).

That lifetime residents of the area have never seen anything similar demonstrates the rarity of this type of pollen transport event. As stratigraphic pollen data often integrate several decades in a single sample, however, such rare events may account for some of the variability often seen in the records of high Arctic regions. Similar rare transport events may account for some of the noise in ice-core records of pollen, dust, charcoal and other particulates and aerosols.

Ian D. Campbell*, **Karen McDonald†**,
Michael D. Flannigan*, **Joanni Kringayark‡**

*Canadian Forest Service, 5320-122 Street,
Edmonton, Alberta, T6H 3S5, Canada
e-mail: icampbel@nrcan.gc.ca

†Environment Canada,
Edmonton, Alberta, T6B 2X3, Canada

‡Government of the Northwest Territories,
P.O. Box 59, Repulse Bay,
Northwest Territories, X0C 0H0, Canada

1. Nichols, H., Kelly, P. M. & Andrews, J. T. *Nature* **273**, 140–142 (1978).
2. Seeley, T. D. et al. *Sci. Am.* **253**, 128–157 (1985).
3. Geurts, M.-A., Lagarec, D. & Guay, F. *Bull. Soc. Bel. Geol.* **93**, 329–333 (1984).
4. Bassett, I. J., Crompton, C. W. & Parmelee, J. A. *An Atlas of Airborne Pollen Grains and Common Fungus Spores of Canada* (Canada Dept of Agriculture, Ottawa, 1978).
5. Faegri, K., Kaland, P. E. & Krzywinski, K. *Textbook of Pollen Analysis* (Wiley, New York, 1989).

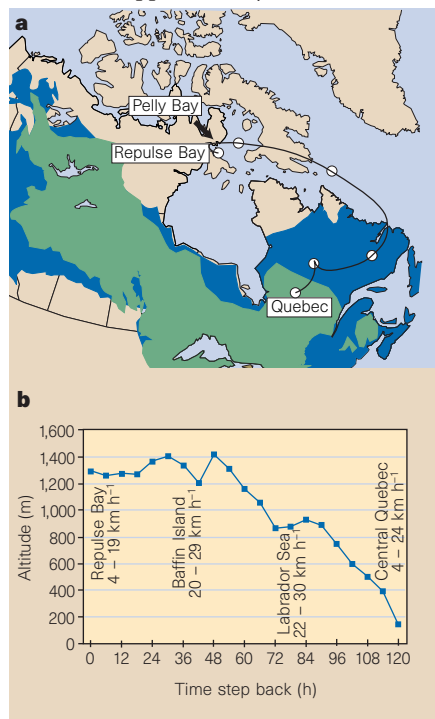


Figure 1 Movement of pollen. Back-trajectory analysis indicates that the pollen was picked up by strong surface winds on 1 June in central Quebec and lifted to and transported at the 850-hPa level to Repulse Bay, where it was deposited in calmer air developed by a low-pressure system on the night of 5 June and early morning of 6 June. **a**, Shading indicates ranges of *Pinus banksiana* (green) and *Picea glauca* (dark blue). The solid line is the 850-hPa trajectory; symbols on the trajectory line mark 24-hour intervals. **b**, Altitude of trajectory from 1 to 6 June.

Cause and effect in evolution

The need to see ‘purpose’ in evolution, or at least some internal drive to help the blind processes of random variation and natural selection, is remarkably resilient¹. Recent manifestations in the scientific literature imagine evolved mechanisms that actively promote further evolution or that facilitate rapid response to changed conditions. For example, Rutherford and Lindquist² (and the authors of related commentaries^{3,4}) suggest that the heat-shock protein Hsp90, by stabilizing developmental pathways, fosters the accumulation of hidden variants that can be exposed by environmental challenges and subsequently fixed by selection.

This is interpreted as “an explicit molecular mechanism that assists the process of evolutionary change”² (or even “a way of saving up mutations for a rainy day”⁴). Similarly, it is widely believed that organisms

increase mutation rates under stressful conditions to improve their chances of hitting on appropriate adaptations⁵.

Such interpretations seem to call for the evolution of properties that anticipate future needs. But selection lacks foresight, and no one has described a plausible way to provide it. In principle, group selection might produce results that seem to escape this limitation. For example, increased mutation rates may indeed allow populations to adapt more quickly to changed conditions, even though they harm most individuals. The evolutionary problem is that such group benefits are usually weaker than individual costs, in a well-defined sense that makes group selection effective only under very restrictive conditions⁶. So, in general, we need explanations that are based on individual fitness differences⁷.

From this perspective, the obvious function of Hsp90 is to prevent abnormalities of the kinds that appear when it is compromised. Up to a few per cent of adults heterozygous for a mutation that inactivates Hsp90 display significant morphological abnormality, so clearly there is selection to maintain its function. Likewise, increased mutation under stress might plausibly arise from trade-offs affecting individual fitness: stressed cells may simply be unable to maintain normal DNA repair without sacrificing other vital functions.

In the natural world, only living things (and their artefacts) have ‘purposes’, and natural selection is the ultimate source of all such ‘purposeful’ design⁸. When speaking of the function or purpose of some feature of an organism, we are therefore referring to the selective advantages that brought the feature into being and that maintain it in the face of recurrent damaging mutations. It is especially important, in any discussion of evolutionary processes, to observe the distinction between function or purpose on the one hand, and effect or consequence on the other. This is not a semantic quibble. Cosmic rays affect evolution by causing mutations, but we would not claim that they exist for that purpose. Similarly, developmental buffering and variable mutation rates may influence the course of evolution, but this does not mean that they evolved to that end.

W. Joe Dickinson, Jon Seger

Department of Biology, University of Utah,
Salt Lake City, Utah 84112-0840, USA
e-mail: dickinson@bioscience.utah.edu

1. Dennett, D. C. *Darwin's Dangerous Idea* (Simon & Schuster, New York, 1995).
2. Rutherford, S. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
3. Cossins, A. *Nature* **396**, 309–310 (1998).
4. Pennisi, E. *Science* **282**, 1796 (1998).
5. Pennisi, E. *Science* **281**, 1131–1134 (1998).
6. Williams, G. C. *Adaptation and Natural Selection* (Princeton Univ. Press, NJ, 1966).
7. Bell, G. *Selection: The Mechanism of Evolution* (Chapman & Hall, New York, 1997).
8. Dawkins, R. *The Blind Watchmaker* (Norton, New York, 1986).

From an old world order to show 'n' sell

Museums and American Intellectual Life, 1876–1926

by Steven Conn

University of Chicago Press: 1998. 293 pp.
\$32.50, £25.95

Thomas F. Gieryn

These days, visitors to museums expect to be entertained, or maybe inspired. A Moon rock or a Matisse masterpiece is displayed so as to evoke wonder. The whiz-bang, hands-on, interactive exhibits through which “children of all ages” can (re)learn principles of centripetal and centrifugal force are there mainly to be fun, and if those who play with them happen to learn a little physics along the way, fine.

It was not always so, as historian Steven Conn tells us in this study of American museums between 1876 and 1926. Then, museums were sites of knowledge-making, and the objects inside — systematically coded, classified, labelled and displayed in “endless glass cases” — were assumed to convey meaning all by themselves. Conn’s mission is to describe how museums of that period functioned as places of intellectual and scientific work, and how, gradually, that role passed to other sorts of institutions.

The diversity of its case studies sets the book apart from the recent flurry of academic analyses of museums. Although limited to museums in the United States between its centennial and sesquicentennial celebrations (each important for the changing form and function of the museum), Conn roams around the many kinds of museums then being built: natural historical, ethnological, commercial, historical and art museums each get a chapter.

This is a good strategy, because it allows Conn to show how the boundaries among these different ontological domains were drawn, contested and redrawn by museum directors seeking to decide which objects were within their purview — and which properly belonged elsewhere. A stone tool from some ‘primitive’ African or South American tribe could, for example, have easily ended up in any of the museums that Conn chooses for in-depth study. Most of these happen to be in Philadelphia: the Academy of Natural Sciences, the University Museum (a centre of ethnological research at the University of Pennsylvania), the Philadelphia Commercial Museum, the historical collection of Henry Mercer in suburban Bucks County and the Pennsylvania Museum (which would become the Philadelphia Museum of Art).

These museums were united in their commitment to an “object-based epistemology”, which Conn describes as the belief that

material objects carry their own knowledge, and that new interpretations could be constructed visually from the inspection of artefacts. However, objects could not become founts of novel understandings if they were randomly gathered or displayed, and neither is there any epistemological payoff from the merely bizarre or freakish (mainstays of an earlier generation of museums). Rather, objects become the source of scientific knowledge only when they are systematically classified and arranged.

The prevalent frame for such categorizations was an evolutionary theory of progress that displayed humankind as moving from savage, to barbaric, to civilized. In effect, each artefact on display stood metonymically for the knowledge that American (and European) culture was the acme of technological prowess, commercial savvy and artistic skill. At the University Museum, for example, this hierarchy of evolutionary progress is architectural: from an entrance landing, visitors either go upstairs to find the genuine civilizations of the Mediterranean and the Near East, or downstairs to the primitive relics of aboriginal peoples of North America and Oceania.

The case for the centrality of an object-based epistemology to the success of these museums is most persuasively made for natural history museums. The fossil specimens assembled at the Field Museum in Chicago



Seat of knowledge: but, like many others, the University of Pennsylvania Museum saw its clientele change from the expert researcher and matriculating student to the turnstile-spinning masses.

were essential for cutting-edge developments in palaeontology, just as dinosaur bones there and elsewhere took centre-stage in debates over Darwinism. But the argument is overdrawn in the case of commercial museums, where the production of new scientific understandings seems incidental to their other missions. The Commercial Museum of Philadelphia gathered mountains of manufactured goods, not to inspire anything epistemological, but to encourage American business — with the “obvious” superiority of their products on display — to expand their markets and investments to foreign shores. The objective seems ideological instead of cognitive: the Commercial Museum sought to undermine economic isolationism while creating the possibility that the United States could become a global power through trade (avoiding messy military interventions or colonial empires).

In fact, Conn’s discussion of why and how museums ceased to be centres of intellectual or scientific work is more interesting and convincing than his discussion of how they got that way. The author would like us to believe that museums lost their knowledge-making role because the object-centred epistemology itself lost currency: objects, specimens and artefacts could simply no longer say that much about the biological or historical processes that interested experts. However, Conn discusses several other coincident developments that might actually be more consequential for the transformation of museums by the 1920s than some epistemological sea-change. For example, especially in America, universities and colleges grew to become the centres of research that they are today: New York’s American Museum of Natural History was once the Mecca of anthropological research, but not after Franz Boas moved its epicentre uptown to Columbia University in 1905. Also, changes in scientific methodology put museums out of the loop.

W. WITTE (LOWER PHOTO)/UNIV. PENNSYLVANIA MUS., ARCHAEOLOGICAL ANTHROPOLOGY

The rise of experimentalism in biology and the rise of hermeneutic studies of historical texts made natural history collections and dusty quotidian artefacts almost irrelevant to the production of new knowledge. Finally, museums struggled to define their preferred clientele, and ultimately the expert researcher lost out. Without matriculating students whose university tuition would cover the costs of laboratories and libraries, museums increasingly played to the masses (and, of course, to wealthy patrons). There was little revenue to be generated by the research functions of museums, but spinning turnstiles and donations of priceless artefacts from aristocratic donors kept the tills full.

Or maybe museums ceased to be centres of intellectual work because Conn's object-based epistemology was never really based on objects at all. Perhaps meaning did not reside in the artefact itself, but only artefacts embedded in a story of progress and inevitable human betterment. As that myth lost credibility, first with the Great Depression and then a second world war, museums gave up their intellectual aspirations in favour of entertainment and awe. □

Thomas F. Gieryn is in the Department of Sociology, Indiana University, Bloomington, Indiana 47405, USA.

From entropy to Duino

Ludwig Boltzmann: The Man Who Trusted Atoms

by Carlo Cercignani

Oxford University Press: 1998. 329 pp.
\$49.95, £29.50

G. F. Bignami

Experimental observations in microphysics date back to the discovery of the electron in 1897, little more than 100 years ago. At the turn of the century, while coming to understand the elementary structure of matter, natural science needed to bridge the gap between micro and macro: to realize that the macroscopic phenomena apparent to our senses are governed by microphysics, through the motions and interactions of vast quantities of particles.

Ludwig Boltzmann (1844–1906) is the scientist to whom, more than anyone else, we owe the great conceptual leap that yielded a joint view of mechanics, statistics, radiation and thermodynamics. He had the courage to abandon avenues of thought, even deeply entrenched ones, when he knew, or felt, that they had been overtaken by reality.

Abandoning ingrained, reassuring habits is frequently the mark of the great scientist and philosopher. Boltzmann had that mark, on both counts, showing, for example, the limitations of the philosophy of Ernst Mach,



A man of principle: Boltzmann's equation is a permanent testament to this great scientist's vision.

an influential and inveterate enemy of statistical mechanics, and paving the way to Max Planck and his quantum jump.

Boltzmann paid dearly for his vision, in terms of his own psychological stability and mental health, up to his tragic suicide. He was embedded in the 'certainty crisis' which characterized the *weltanschauung*, or philosophy of life, of intellectual Vienna at the turn of the century. Take Boltzmann's statement on the nature of the Universe: "a sea of generalized uniformity, with here and there some temporary statistical ripples representing individual worlds." Elegant and accurate enough today, but of a shocking and destabilizing depth of insight coming, as it did, decades before modern cosmology.

Boltzmann had a very intense life, dominated by insecurity, academic feuds and a passion for science. He was also a sensitive musician and a lover of nature, with a keen eye for picking out and explaining to his children, for example, evolutionary mechanisms. He would tell them long tales of a great English scientist sailing around the world in a ship called *Beagle*, and from that explain why foxes have fur, or birds feathers. His loving wife Henriette once asked him if their love was also explained by Darwin's theory. It was she who suggested a holiday when she felt that his mental state was slipping. But Boltzmann was not to return from Duino, then on the Austrian Adriatic coast.

Many of the details of Boltzmann's life and thought were lost when documents and manuscripts went up in flames during the Second World War. Moreover, what's left of his work is in a secret shorthand, under-

standable only by his granddaughter, Ilse Fasol-Boltzmann, editor of the authoritative *Ludwig Boltzmann: Prinzipien der Naturphilosophie* (Springer, 1990).

Cercignani's beautiful book has the merit, first of all, of bringing Boltzmann fully back to life, as a scientist, a philosopher and a poet, thanks to painstaking research. Not only do we feel the depth of the man who gave us the H-theorem, the time-arrow and the probabilistic interpretation of entropy ($S = k \log W$, as engraved on his tombstone, or what Einstein called the Boltzmann principle). We are also made to appreciate the sensitive mind that could write the 100-line jocular poem "Beethoven im Himmel", in which angels tell Boltzmann's soul on its way to heaven that the song they sing was composed by Beethoven "upon the Lord's command", and is by far the best in their repertoire.

We also meet Boltzmann's *Populäre Schriften* — "popular writings" — containing lasting, if non-technical, contributions to culture. Among these is the first lecture he gave on a natural philosophy course, when he took up Mach's former position at the University of Vienna. The largest hall had been chosen, but the event had been so extensively publicized in the press that more than 600 people crowded in, some of them having to stand. It was a resounding success, in spite of all his previous insecurity and apprehension, and Emperor Franz Joseph asked him to Court for a personal word of congratulation.

Two years later, Boltzmann visited the United States for the last time, and wrote a sparkling account of his travels in *Reise eines deutschen Professors ins Eldorado* — Eldorado being the great American West. Here we learn, among other things, that wine was hard to find in 1905 California, and that, in naively asking for wine, Boltzmann caused as much embarrassment as if he had asked for the address of a house of ill repute.

Coming as it does from a leading authority on the mathematical theory of Boltzmann's equation (and much more), Cercignani's book gives a complete and very impressive account of Boltzmann's work. The going is frequently *allegro con brio*, at times uphill and strenuous, but always fascinating. This book brims with the insight and passion of a lifetime's work on Boltzmann, expanded over the years to encompass a consistent view of European science in Boltzmann's times. (The 16 appendices can stand alone as a textbook for a course in statistical mechanics "from a historical but rigorous standpoint".)

A great figure in science is best etched, and most easily understood, by comparison with his contemporaries. Cercignani does this superbly, mastering the figures of James Maxwell, Hendrik Lorentz, Planck and others, and their influence on Boltzmann and his on them. In his foreword, Roger Penrose sums it up by noting that Boltzmann stands

as a link between Maxwell and Einstein, connecting the physics of two centuries.

In the end, however, something remains beyond reach, even for Cercignani. "We can conjecture, but we cannot present conjectures as proved circumstances. Thus, we must leave a veil of mystery over the final act of the life of Ludwig Boltzmann." □

G. F. Bignami is at the Italian Space Agency and University of Pavia, Via E. Bassini 15, 20133 Milan, Italy.

Aiming for that goal of zero emission

Green Chemistry: Theory and Practice

by P. T. Anastas and J. C. Warner
Oxford University Press: 1998. 134 pp.
£45, \$85

Green Chemistry: Frontiers in Benign Chemical Synthesis and Processes

edited by P. T. Anastas and T. C. Williamson
Oxford University Press: 1998. 364 pp.
£65, \$115

Roger Sheldon

In 1856, W. H. Perkin, who was 18 years old at the time, attempted to synthesize quinine by dichromate oxidation of a raw material obtained from distilled coal tar. This resulted in the serendipitous discovery of the first synthetic dye, aniline purple. Within a few years a commercial plant was in operation, and this is generally recognized as the first industrial organic synthesis. The organic chemical industry was born, and the next 100 years witnessed the remarkable contributions of synthetic organic chemistry in increasing life expectancy, providing food, clothing and shelter and improving the quality of life in general.

The turning point in this remarkable success story came in the early 1960s, with the realization that a price was being paid for these achievements. The publication of Rachel Carson's *Silent Spring* in 1962 alerted the public to the negative effects of the accumulation of synthetic chemicals, such as DDT, in the environment. The environmental movement was born, and for the next three decades much effort was devoted to studying the effects of various synthetic chemicals on the environment, a recent example being the role of chlorofluorocarbons in the depletion of atmospheric ozone. The emphasis was clearly on the products rather than on the processes by which they were being produced.

In the 1990s, we are now focusing on the processes used in synthesizing organic chemicals. And the conclusion is clear: many of these processes generate enormous waste, for example as inorganic salts in aqueous

effluents and volatile organic molecules in air emissions. The cause is also clear: many industrial organic syntheses use antiquated technologies involving stoichiometric inorganic reagents such as the dichromate oxidation used by Perkin almost a century and a half ago. Historically, as Paul Anastas and John Warner point out in *Green Chemistry: Theory and Practice*, synthetic chemists have not been particularly environmentally conscious, since their involvement was at the beginning of the chemical synthetic chain whereas problems were mostly encountered at its end.

The solution is the replacement of these technologies with cleaner catalytic alternatives. The emphasis is on eliminating waste at source — primary pollution prevention — rather than finding incremental end-of-pipe solutions. This has now become known as green chemistry, and is defined by Anastas and Warner as: "The utilization of a set of principles that reduces or eliminates the use or generation of hazardous substances in the design, manufacture and application of chemical products". The tools of green chemistry are alternative feedstocks, solvents and reagents, and catalytic versus stoichiometric processes. And the book introduces and explains the concept of atom efficiency in organic synthesis — fundamental to pollution prevention. (However, I found the treatment rather superficial, and was surprised to read that light is considered a feedstock.)

Hence, the concept of green chemistry encompasses both alternative products and alternative processes. It is a useful working definition, but concentrates too much, for my taste, on hazardous substances rather than the goal of reducing all waste (zero-emission plants).

The book defines the 12 principles of green chemistry — described as the Hippocratic oath for chemists. The first principle is that it is better to prevent waste than to treat or clean up waste after it is formed; the other 11 can be roughly paraphrased as: processes should be atom- and energy-efficient, use renewable rather than depleting raw materials and avoid using toxic and/or hazardous reagents and solvents, and products should be designed for non-persistence in the environment. The concept of green chemistry is closely related to, or perhaps even synonymous with, that of sustainability (defined as the ability to maintain the development of the quality of life while not compromising the ability of our progeny to do the same).

The examples of green chemistry used are predominantly taken from the Green Chemistry Challenge Awards, which are sponsored by the US Environmental Protection Agency and, consequently, are heavily biased towards the United States. And some are rather contrived — a broad scope is implied for a photochemical alternative to Friedel Crafts acylation, but it is actually very limited.

Anastas and Warner's prediction of future trends is a mixed bag of well-defined goals, for example new catalytic oxidation technologies with O₂ or H₂O₂, and rather nebulous and/or contrived concepts such as "biomimetic multifunctional reagents" and "combinatorial green chemistry" (whatever that may be).

Green Chemistry: Frontiers in Benign Synthesis and Processes, edited by Anastas and Tracey Williamson, his colleague at the Environmental Protection Agency, contains more chemistry for chemists to get their teeth into. It is a useful collection of articles devoted to different aspects of green chemistry and is complementary to the Anastas/Warner book. One chapter (and the foreword) is written by Barry Trost, the champion of atom efficiency in organic synthesis.

Again, however, the contributions are overwhelmingly biased towards US academia and, as a result, the examples tend to be rather dry and concentrate on the authors' own work, thereby failing to provide either an overview or industrial relevance for the topics. It is also a pity that there are not more contributions from industrial groups — Monsanto, Mitsubishi, Rhône-Poulenc and Lonza come readily to mind. □

Roger Sheldon is in the Department of Organic Chemistry and Catalysis, Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands.

Information, information

The Touchstone of Life

by Werner R. Loewenstein
Oxford University Press: 1999. 366 pp. \$30
Nancy Lane

Information — biomolecular information — is this book's central character, and the story tells how living systems have evolved. According to Werner Loewenstein, evolution is guided by "the principle of information economy in self-developing systems, by an accretion of information circles in a process of least information cost". To grasp this argument, one must read on, crossing a terrain of historical detail and complex metaphor.

Starting with the Big Bang and the origin of the Universe, Loewenstein sets up a battle between two information systems, with information-conserving connections (information loops and self-catalysing systems) winning the day. The text, peppered with intriguing facts and often speculative figures, then proceeds to the complexities of modern cellular biology. A major theme is cell communication. The specificity of biological control and its strict information economy are also stressed throughout. Pos-

Science in culture

Babes in bottles

The anatomical art of Frederik Ruysch
Martin Kemp

The sight of pickled babies and their severed parts, floating in jars of balsamic liquor, adorned with lace cuffs, elaborate hats and festoons of beads, is bound to make the modern spectator feel uneasy. Yet seventeenth-century visitors to the "cabinet" of Frederik Ruysch in Amsterdam appear to have expressed no disgust. The science and aesthetics of anatomical preparations were clearly very different from those in a modern anatomical museum.

Dr Ruysch was Praelector Chirurgiae et Anatomiae for the City of Amsterdam over a remarkable span of time, from 1667 until his death in 1731, as well as serving as the supervisor of its botanical garden. He was the subject in 1670 and 1683 of two paintings in the famous series of Dutch "anatomy lessons", of which Rembrandt's recently restored "Anatomy of Dr. Nicolaes Tulp" is the most celebrated. An acclaimed teacher and popular public dissector, Ruysch gained international fame for anatomical preparations made by his secret techniques of injection, in the form of isolated systems such as blood vessels and of whole corpses prior to dissection.

One witness to Ruysch's collection delighted in "groves of plants and designs of shell-work with skeletons, and dismember'd limbs ... with apposite inscriptions from the Latin poets". His grand house in Amsterdam became a tourist attraction, visited by "generals of armies, ambassadors, electors, and even princes and kings". One such monarch, Tsar Peter the Great of Russia, was so impressed that he purchased the whole ensemble in 1717 for the princely sum of 30,000 guilders and installed it in his new city of St Petersburg, where many specimens still survive in the Kunstkammer of the National Academy of Sciences.

One lost exhibit, illustrated in a folding plate in Ruysch's *Thesaurus Anatomicus* of 1717, took the form of a small mountain scattered with gall, kidney and bladder stones on which flourishes a forest of injected vessels. The elaborate tableau is completed with three grieving infants' skeletons, creating a miniature 'theatre' on the theme of the transience of life. The "trees" of vessels and bodily rocks stress the microcosmic conception of the human body as a "lesser world".

Many of the preparations displayed the

sible mechanisms for the cellular basis of cancer, theories of neuronal communication and, especially, the thorny issue of the basis of human consciousness are all dealt with in some detail, albeit, unsurprisingly, inconclusively.

Alas, the story's complicated grandeur will probably be incomprehensible to anyone not familiar with biophysics and molec-



Ruysch's "Head of Infant with Dissected Cranium and Glass Eyes", from the Kunstkammer of Peter the Great in the Museum of Anthropology and Ethnography in St Petersburg.

whole or parts of babies, whose small corpses were available to him through his obstetric activities and were readily susceptible to injection. Ruysch did not concentrate on the kind of monstrous deformities that attracted popular curiosity, but focused instead on preparations that aspired to evoke the beauty of "innocents", dead before their time. In the Kunstkammer, we see amputated limbs emerging from lace cuffs, tenderly supporting a dissected heart, an intact placenta. An eye dangles from tiny fingers on a fine thread in an example remaining in Leiden. In the above example from St Petersburg, now housed and bottled far more plainly than in its original context, an infant's head, horizontally sectioned to reveal the open cranium, conceals its severed neck behind a fine scarf of lace, while its glass eyes regard the enquiring viewer with an eerie similitude of consciousness.

Having experienced fleeting lives, at best, Ruysch's infant subjects have been accorded a kind of immortality, cheating rigor mortis and even long-term decomposition. At least, this is how they were seen at the time. As a poem printed in Ruysch's *Thesaurus* said, "Through thy art, O Ruysch, a dead infant lives and teaches and, though speechless, still speaks. Even death itself is afraid."

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

ular biology, while those who are will already know it well. For whom, then, is the tale told? This book is not for lay-readers, so perhaps it is for those with a background in the biomolecular sciences who want to be brought up to date, or those who want to brood on our origins.

Loewenstein's extensive musings and explanations are generally far from tedious.

Indeed, he has clearly thought long and hard about how to present complex concepts and what metaphors will be effective. Concepts of "the circus", "loops within loops", "multi-plex" forms of apparatus and driver "demons" enliven the arguments. He weaves his story in an engrossing theatrical context, centring on the conservative forces of "Lady Evolution" that led to the cellular subunits comprising today's organisms, and beyond to cell subordination and differentiation.

There seems little doubt that Loewenstein sees this book as an opportunity to broadcast his contribution to the field of cell-to-cell communication and the control of cell growth. Channels — his consuming passion — are paid particular attention. His bias towards the importance of his own laboratory's work, and his tendency to ignore that of other, particularly non-American, labs is striking. The blurb even claims that Loewenstein is the "man who ... opened the field of cell-cell communication".

Loewenstein plumbs the known depths of cellular and molecular biology with verve, always attempting to set the facts in the context of how the processes might have evolved and why they need to be so sophisticated and complex. In all this he succeeds in gripping fashion, yet the demands he makes of his readers may well deter all but the highly knowledgeable or utterly determined. □

Nancy Lane is in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 2 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1997 and issued at least four separate numbers by the end of May 1999.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Isobel Flanagan, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4542. e-mail: i.flanagan@nature.com

Energy traps in atomic nuclei

Philip Walker & George Dracoulis

A small proportion of atomic nuclei can form highly excited metastable states, or isomers. Of particular interest is a class of isomers found in deformed axially symmetric nuclei; these isomers are among the longest-lived and have the potential to reach the highest energies. By probing their properties, insights into nuclear structure have been gained. The possibility of stimulated isomer decay may ultimately lead to new forms of energy storage and γ -ray lasers.

Research into the structure of atomic nuclei concentrates largely on the limits of nuclear existence. This is where new physics is most likely to be—at the limits of neutron number (N), proton number (Z), excitation energy (E) and total nuclear angular momentum (I). For example, very neutron-rich nuclei have been shown¹ to have neutron ‘haloes’, favoured nuclei at very high angular momentum become ‘superdeformed’² and there is a continuing quest to find ‘superheavy’ nuclei³, whose predicted ground states may form an island of metastability in the nuclear mass table.

Another kind of metastability is found in the quantum states of nuclei at high excitation energy; such states are named ‘isomers’, by analogy with chemical isomers. Two chemical isomers have the same constituent particles, but in different physical configurations. It is likewise for two nuclear isomers. However, whereas the two chemical isomers would have essentially the same energies, a nuclear isomer can have an excitation energy of several million electron volts (MeV) relative to the ground state.

As isomer de-excitation depends more-or-less entirely on spontaneous emissions (nuclei are isolated objects, so that collisional de-excitation does not occur), isomer half-lives can be remarkably long⁴. An example occurs in the hafnium nucleus ^{178}Hf . It contains 72 protons and 106 neutrons (that is, 178 nucleons), but has no angular momentum (or spin) in its stable ground state because the nucleons couple in spin-zero pairs. When just two of the protons and two of the neutrons recouple, they can make a state of unusually high spin (16, in units of $\hbar$ —Planck’s constant divided by 2π), the half-life of which is 31 years. This isomer is ‘suspended’ at an energy of 2.4 MeV above the ^{178}Hf ground state.

This energy (2.4 MeV) is much higher than found for atomic metastable states, which are usually at energies of a few electron volts (eV) and can de-excite rapidly by collisions with other atoms. However, isomer energies are much lower than the energies released in nuclear fission. The fission of a nucleus such as uranium, on which our sources of nuclear power depend, releases an energy of about⁵ 200 MeV (or 30 pJ). Nevertheless, the capacity of isomers to store energy is significant. Consider the tantalum nucleus ^{180}Ta or, more specifically, $^{180\text{m}}\text{Ta}$, where m denotes the metastable, isomeric state of ^{180}Ta . This is the only nuclear isomer to exist naturally on Earth; it survives, with a half-life of more than 10^{15} years, from before the Earth was formed. Natural tantalum exists mainly as ^{181}Ta , with only one part in 10^4 being $^{180\text{m}}\text{Ta}$, and the excitation energy of $^{180\text{m}}\text{Ta}$ is just 75 keV, relative to the ^{180}Ta ground state. Yet the total isomer energy stored in one cubic centimetre of natural tantalum is 30 kJ, and it would be 300 MJ in one cubic centimetre of pure $^{180\text{m}}\text{Ta}$. However, although $^{180\text{m}}\text{Ta}$ exists naturally, and is quite harmless, an efficient method of releasing its energy has yet to be discovered.

Here we address the reasons for the existence of nuclear isomers, and what limits their existence. In so doing, a subtle interplay unfolds, between collective motion (involving the whole nucleus) and individual nucleon dynamics. We focus on a class of ‘spin trap’ isomers, whose numbers are growing thanks to advances in experimental techniques for discovering them. Such progress not only leads to a better understanding of their nature, but also suggests

that laser-stimulated release of the stored energy of nuclear isomers might soon be achievable.

Shape isomers and spin traps

The nuclear isomers found to date arise in distinctly different ways, as illustrated in Fig. 1. It is difficult for them either to change their shape to match the states to which they are decaying, or to change their spin, or to change their spin orientation relative to an axis of symmetry. These situations depend on the detailed ‘shell’ structure of the neutron and proton orbits in each nucleus. The nuclear shells arise in a similar way to atomic shells. In atoms, full (or closed) shells give extra stability, resulting in inert gases at certain atomic numbers; similarly, strongly bound nuclei occur at ‘magic’ neutron and proton numbers. In between the magic numbers, individual nucleon orbits can strongly influence the nuclear properties.

Shape isomers. These occur when there is a secondary energy minimum at large elongation of the nucleus (the primary energy minimum corresponds to the ground state). The clearest examples are the fission isomers in heavy nuclei⁶, such as ^{242}Am . Here, a 2.2-MeV isomer, elongated such that its major-to-minor axis ratio is about 2:1, fissions with a half-life of 14 ms, the longest half-life yet found for an isomer of this type. For some fission isomers, decay back to the ground state by γ -ray emission competes with fission into two lighter nuclei.

Spin traps. A more common form of isomer is the spin trap, whose existence depends on the difficulty in meeting spin selection rules (which embody conservation of angular momentum). The decay path to lower energy states requires a large change in nuclear spin, and therefore the emission of radiation with a high multipolarity (λ , equivalent to the angular momentum carried by the radiation in units of $\hbar$) to match the spin change. For example, if $^{180\text{m}}\text{Ta}$, with a spin of $I = 9$ (in units of $\hbar$) and an excitation energy of 75 keV, were to decay to the ^{180}Ta , $I = 1$ ground state, it would have to emit radiation with $\lambda = 8$. The extreme improbability of this (see Box 1) is reflected in the empirical half-life lower limit for $^{180\text{m}}\text{Ta}$ of 10^{15} years.

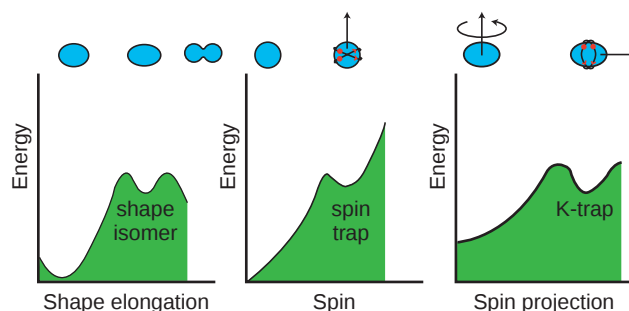


Figure 1 Excitation energy as a function of various nuclear variables. The secondary energy minima are responsible for the different kinds of isomers: **a**, shape isomers; **b**, spin traps; **c**, K-traps. In each case, the relevant nuclear shapes are illustrated; where appropriate, angular momentum vectors are shown as arrows. For both the spin trap and the K-trap, the angular momentum comes from a small number of orbiting nucleons (two are illustrated in red in each case).

Although most spin traps decay by electromagnetic processes (γ -ray emission, or internal conversion—the process by which the de-excitation energy is transferred directly to an atomic electron, expelling it from the atom), examples are known where the decay is mediated by the strong interaction (α -particle or proton emission) or by the weak interaction (β -particle emission or electron capture).

The 1970s and 1980s saw many discoveries of spin traps in spherical or near-spherical nuclei, with MeV excitation energies⁷. An outstanding example is an $I = 34$ isomer⁸ in ^{212}Fr , with a half-life of 24 μs and an energy of 8.5 MeV. Apart from the question of whether more extreme isomers might exist in heavy nuclei such as this, their accessibility is limited by the very reaction process that is needed to make them. The fusion of colliding nuclei is required (see the section ‘Making spin traps’) but after fusion, such heavy nuclei are likely to fission before forming isomers.

K-traps. The third type of isomer, known as the ‘K-trap’, is a form of spin trap whose existence depends not only on the magnitude of the nuclear spin vector, but also on its orientation. K is a quantum number that represents the projection of the total nuclear spin along the symmetry axis of the nucleus—this type of isomer arises only in axially symmetric, deformed nuclei that are found (in N, Z space) well away from the closed shells that favour spherical shapes. The nuclei that form K-traps are all prolate, with a shape rather like that of a rugby football (or Australian Rules or American football), where the long axis is the axis of symmetry.

A strict selection rule would require that the multipolarity of the decay radiation, λ , be at least as large as the change in the K -value ($\lambda \geq \Delta K$). However, symmetry-breaking processes make possible transitions that violate the K -selection rule; such ‘K-forbidden’ transitions are hindered, rather than literally forbidden⁹. For example, in $^{180}\text{Hf}_{108}$ (with 72 protons and 108 neutrons) there is an $I = 8$, $K = 8$ isomer at 1.1 MeV with a half-life of 5.5 hours. Its most probable decay is by a 58-keV, $\lambda = 1$ transition to an $I = 8$, $K = 0$ state, so that $\Delta K = 8$, clearly violating the rule that $\lambda \geq \Delta K$. (In principle, a K -allowed path is available, via a 1.1-MeV, $\lambda = 8$ γ -ray to the $I = 0$, $K = 0$ ground state, but the high multipolarity results in a transition rate that is not competitive).

The mechanisms leading to nuclear isomerism—shape changes, spin trapping and K -trapping—can reinforce each other. Indeed, one of the regions of the nuclear mass table with the most isomers, around ^{180}Hf , depends on both spin and K -trapping. The occurrence^{4,10,11} of highly excited (>1 MeV), long-lived (half-life >1 ms) isomers as a function of neutron and proton number is

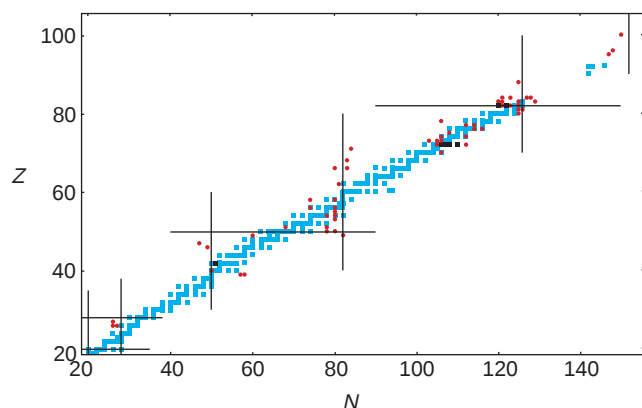


Figure 2 Chart of atomic nuclei, as a function of neutron number (N) and proton number (Z). Naturally occurring isotopes are represented by blue squares. The straight lines are drawn for the ‘magic numbers’ that correspond to closed shells (and spherical shapes). The other symbols indicate isomers with excitation energies greater than 1 MeV and long half-lives: red circles for >1 ms, and black squares for >1 hour. The data are taken from refs. 4, 10, 11.

shown in Fig. 2. The $N = 108$, $Z = 72$ mass-180 region of nuclei is seen to contain the largest number of very long-lived (half-life >1 hour) isomers.

However, what makes the mass-180 isomer region particularly interesting is that there may be in this region many long-lived isomers just waiting to be revealed, some of which could be at exceptionally high excitation energies. The ideas that lead to this optimistic viewpoint will be explained below.

Making spin traps

Firing a bullet into a punchbag can make the latter locally hot, but it will not move very much. The lightweight bullet may have high kinetic energy ($0.5MV^2$), but only low momentum (MV) and low angular momentum (MVr), even if it hits the punchbag well off-centre (here M is the mass of the bullet, V is its velocity, and r is the impact parameter). It is the same with nuclei. Firing a proton into $^{180}\text{Hf}_{108}$ with enough energy to overcome the electrostatic repulsion makes $^{181}\text{Ta}_{108}$ at relatively low angular momentum, but at sufficiently high excitation energy to ‘boil off’ two neutrons to become $^{179}\text{Ta}_{106}$. γ -ray emissions follow, corresponding to transitions between quantum states, in a sequence that takes ^{179}Ta down to its ground state, but only through states with low spin.

Fusion reactions. Heavier projectiles are needed to make high-spin isomers, with fusion-evaporation reactions of the type illustrated in Fig. 3a. The record for a K-trap has been set¹² by colliding ^{48}Ca into ^{130}Te . After fusion, which forms ^{178}Hf , the excitation energy favours the emission of four neutrons which would lead to states in ^{174}Hf . Occasionally only three neutrons are ‘evaporated’, resulting in a rapidly spinning ^{175}Hf nucleus, which sometimes is trapped in an isomeric state with $(57/2) \hbar$ of spin at an excitation energy of 7.5 MeV.

However, the low probability of the formation path makes difficult the study of ^{175}Hf at high spin with this reaction. (As a result, the 7.5-MeV isomer has yet to have its half-life measured.) To do better, an initial beam or target with more neutrons is needed, but for calcium and tellurium these heavier isotopes are all unstable, with half-lives shorter than days, so they are unavailable for beams and targets at present. Developments in radioactive-beam technology¹³ should soon remedy this.

The heaviest isotope of hafnium that presently can be studied with fusion-evaporation reactions is $^{178}\text{Hf}_{106}$, made by colliding ^4He into ^{176}Yb , and involving the evaporation of two neutrons. This reaction brings in, at the most, 16 units of spin, just enough¹⁴ to make the $K = 16$ trap at 2.4 MeV, which has a half-life of 31 years.

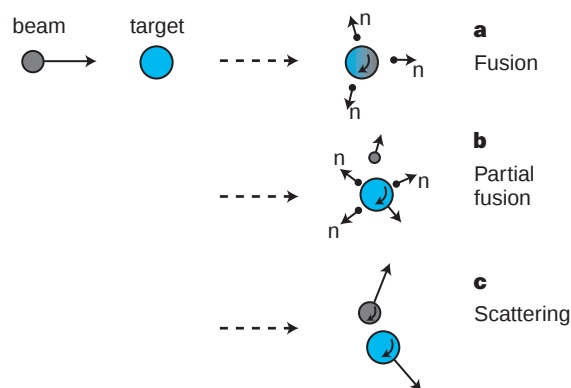


Figure 3 Some of the possible results of a beam (projectile) nucleus colliding with a target nucleus. **a**, Fusion, together with the evaporation of some neutrons (a process that takes about 10^{-19} s); **b**, partial fusion, again with some neutron evaporation; **c**, scattering (with or without some nucleons being exchanged). In each case, specific nuclei are produced at excitation energies up to ~ 10 MeV, energy that they subsequently shed by γ -ray emissions.

This half-life is long enough to enable sufficient quantities of the isomer to be accumulated for secondary experiments^{15,16}. Other, higher-spin isomers are predicted^{17–19} to exist in ^{178}Hf and in heavier hafnium isotopes. However, they are out of experimental reach with fusion-evaporation reactions, and other approaches are needed.

Partial fusion reactions. Partial fusion, or projectile breakup, reactions (Fig. 3b) can impart a few more units of angular momentum in some cases. For example, ^{178}Hf can be studied by bombarding ^{176}Yb with ^9Be . Sometimes, the latter breaks apart in the field of the former, and an α -particle (^4He) flies off, while the remaining fragment—unstable ^5He —fuses with the target. The simultaneous detection of γ -rays and the tell-tale α -particle distinguish the hafnium nuclei from the more copious fusion products (various tungsten isotopes). After the evaporation of three neutrons, the residual ^{178}Hf nuclei can have up to 23 units of spin, which is enough to reach²⁰ several rotational excitations on top of the 2.4-MeV, 31-year trap. The information contained in the properties of rotational states such as these will be discussed in the next section ('Spin-trap rotations').

Scattering reactions. A scattering method (Fig. 3c) has recently led to the identification^{10,21} of several new K -traps. The successful technique involves bombarding heavy, deformed, target nuclei, such as ^{180}Hf (the heaviest stable isotope of hafnium) with heavy, deformed, projectile nuclei, such as ^{238}U (the heaviest naturally occurring nuclide). Even when the beam energy is sufficient to overcome the electrostatic repulsion, the target and projectile nuclei will not fuse because of their substantial positive charge, which immediately causes them to separate. But during their brief encounter, considerable spin can be imparted if the relative orientations at impact are right, rather like colliding two rugby footballs.

This process can make spin traps that are otherwise out of reach, the most strongly produced traps being in the target nuclei themselves¹⁰, with spins in excess of 20 units²². During the nuclear encounter, nucleons can also be exchanged, so that states in nuclei a few neutrons or protons away from the target or projectile may also be reached^{21,22,23}.

Fragmentation reactions. A more radical method involves the use of fragmentation reactions. At relativistic velocities, nuclear collisions are extremely violent, and a vast range of fragments can emerge. The reaction is not selective in terms of its products, but the high velocity of each fragment of the projectile, flying forward after the collision, enables the use of time-of-flight and energy-loss techniques to label its mass and charge. Hence, each product

nucleus, in its ground or isomeric state, can be identified. Though now proven^{24,25} as a method for making and identifying new isotopes and isomers, the method has yet to be applied to searching for new K -isomers in neutron-rich isotopes.

Using these new methods, a large number of new K -traps will probably soon be discovered in stable and neutron-rich nuclei in the mass-180 region.

Spin-trap rotations

The observed properties of nuclei are determined not only by the behaviour of individual nucleons, but also by collective, 'whole-nucleus' effects. For example, a deformed, axially symmetric nucleus can rotate as a whole about an axis that is perpendicular to the symmetry axis. There is no collective rotation about the nuclear symmetry axis because this would result in a system that is indistinguishable from the original. This is why K , the projection of the resultant angular momentum onto the symmetry axis, is a conserved quantity. Non-zero values of K arise from the summed contribution of individual orbiting nucleons, which do not themselves have axial symmetry (Fig. 1c).

The way that a deformed nucleus rotates as a whole continues to challenge our understanding of basic nuclear structure. An important part of the early development²⁶ of the collective nuclear model (which embodies co-operative effects between nucleons) depended on K -traps. Notably, the decay of the $K = 8$ isomer²⁷ in ^{180}Hf gave access to the ground-state rotational band, which consists of a series of collectively rotating states of differing angular momenta (I) for a given value of K (zero in the present case). The energies of these rotor states were observed to have energies proportional to $I(I + 1)$, with a constant of proportionality equal to $\hbar^2/2J$, where J is the moment of inertia.

It was immediately evident that the moment of inertia was less, by a factor of three, than that expected for simple rotation of a whole nucleus. This can be explained if pairing correlations between nucleons exist, leading to superfluidity and implying rotation like a superfluid, rather than a rigid body (Fig. 4). Pairing has a profound effect on nuclei, and it is now well accepted that the superfluid state is the natural state for nuclei in their ground-state configurations. Only in the absence of pairing correlations would a rotating fermion system be expected to display a rigid-body moment of inertia²⁸, and then only on average.

Correlated fermion motion is also seen in superconducting metals at low temperatures, in the form of electron–electron

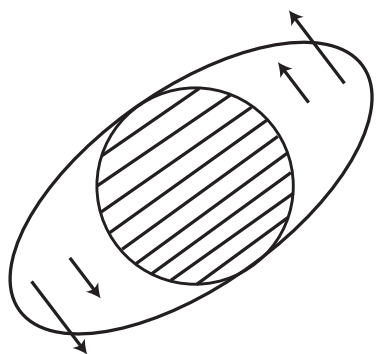


Figure 4 Illustration of the nuclear velocity fields for a deformed, rotating nucleus. We assume the two-fluid model of Rowe²⁹. The arrows represent the nucleon local-average velocity vectors. The spherical superfluid core (hatched region) does not rotate, resulting in a moment of inertia that is substantially smaller than that of a rigid body.

Box 1: Isomer half-lives and widths

There is no strict definition of the minimum half-life that an excited nuclear state needs to possess in order to qualify as an isomer. The qualitative requirement is that the half-life should be long compared with most other excited states, which is generally taken to mean that the half-life should be longer than 1 ns. Consequently, the energy width of the state, Γ (which, by Heisenberg's uncertainty principle, is inversely proportional to the half-life), is narrow—it stands out from the crowd of other states. For example, a state with a half-life of 1 ps has a width of ~ 0.5 meV, while an isomer with a half-life of 1 μs has a width a million times less (0.5 neV).

The γ -ray decay rates are governed by electromagnetic processes. A simple formulation of 'single-particle' transition rates (in which nucleon–nucleon interactions are effectively ignored) was evaluated by Weisskopf²⁸, and these still form a useful benchmark with which to compare measured transition rates. The Weisskopf estimates predict that the transition rate varies as the $2\lambda + 1$ power of the transition energy (where λ is the multipolarity of the decay radiation), and that higher transition multipolarities have lower transition rates. Comparisons with experimental rates enable a variety of nuclear structure effects to be picked out. For example, K -forbidden transitions are systematically slower than the Weisskopf estimates, by orders of magnitude.

pairing. Bardeen–Cooper–Schrieffer (BCS) theory²⁹, which describes this behaviour, was extended³⁰ early on to the study of nuclear superfluids. Whether it is electron–electron pairing in metals that is considered, or nucleon–nucleon pairing in nuclei, the disappearance of resistance to motion at low temperatures has a significant effect on the properties that we observe (and, in the case of superconductors, has important consequences for technology).

Superconductivity in metals can be quenched (destroyed) by heat and/or magnetic fields. In nuclei, the respective analogues are internal excitation energy and collective rotation. Both of these break up nucleon pairs, blocking the overall pairing correlations and destroying the superfluidity. However, nuclei and metals differ in a number of ways, including the small size of the former and their finite particle (nucleon) number, which lead to phase transitions that are not sharp. Although this makes the nuclear superfluid-to-normal phase transition difficult to pin down, it does allow it to be investigated step-by-step, with individual quantum states and the effects of specific nucleon orbits open to observation.

The quenching of nuclear pairing correlations by rapid collective rotation has been studied thoroughly, if indirectly, by interpreting the properties of rotational transitions³¹ at high angular momentum. The process is not fully understood, particularly as the effects of internal excitation energy complicate the picture. In rapidly rotating nuclei, Coriolis forces (resulting from the Coriolis effect that creates cyclonic and anticyclonic wind patterns on the rotating Earth) introduce further complications; they perturb the individual nucleon orbits by creating a preferred rotational direction, which tends to destroy the pairing correlations.

However, by studying slowly rotating *K*-traps, in which the pair-breaking effects due to collective rotation are small, it becomes possible to isolate the effects of internal energy. Each *K*-trap represents an independent configuration (with its own characteristic states of collective motion), and its metastability implies that the nucleon orbits are well-defined. *K*-trap observables, such as the spin quantum number, *I* (resulting from the broken nucleon pairs) can therefore be used to specify which pairs are broken and are thus quenching the superfluidity; and, by measuring the energies of the slowly rotating states that are specifically associated with a given *K*-trap, the moment of inertia can also be deduced.

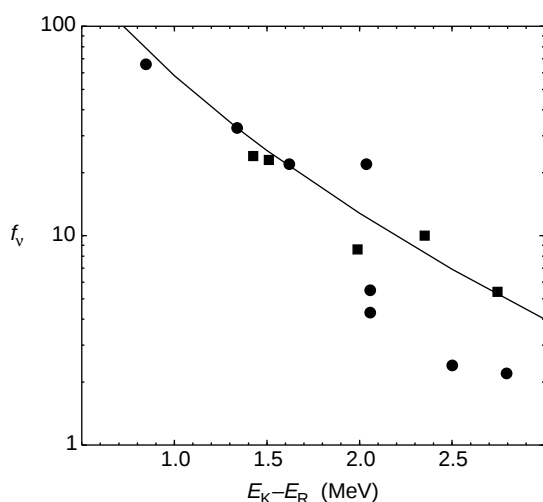


Figure 5 Reduced hindrance as a function of isomer excitation energy, relative to a rotor with the same angular momentum, for selected *K*-forbidden transitions. The solid curve represents a statistical-model estimate based on the density of states. The figure is adapted from ref. 45. Circles represent four-quasiparticle isomer decays in even-*N*, even-*Z* nuclei; squares represent five-quasiparticle isomer decays in odd-mass nuclei (a pairing energy of 0.9 MeV has been added to the excitation energy above the ground state).

From an experimental viewpoint, the rotating *K*-trap with the most unpaired nucleons observed to date³² is in ¹⁷⁸W, in which an eight-quasiparticle, 6.6-MeV isomer has been identified, with four unpaired neutrons and four unpaired protons. (The term ‘quasi-particle’ is used because the pairing interactions that exist between the remaining paired nucleons modify the overall wavefunction and excitation energy. The number of quasiparticles is equal to the number of unpaired nucleons.) Despite the sizeable number of broken pairs, the moment of inertia obtained from the eight-quasiparticle rotational band is still only 60% of the rigid-body value. This compares with 34% for the ¹⁷⁸W ground-state (fully paired) rotational band.

A straightforward conclusion (from the less-than-rigid moment of inertia), supported by detailed analysis^{33,34}, would be that pairing correlations persist even when there are four unpaired nucleons of each type, albeit substantially less strongly than in the ground state. However, alternative interpretations are possible. It was pointed out some time ago³⁵ that the nuclear shell structure may significantly modulate the average fermionic behaviour. Detailed rotor calculations, which allow for the generation of angular momentum both along the symmetry axis (non-collective), and perpendicular to it (collective), are able to reproduce³⁶ the observed moments of inertia without invoking pairing correlations. In other words, the lower than rigid-body value (60%), found for the ¹⁷⁸W eight-quasiparticle rotational-band moment of inertia, can be interpreted as a residual shell (or nucleon orbit) effect, and still be compatible with the complete loss of pairing correlations. But the model calculations are not yet on a sufficiently firm theoretical basis for a satisfactory conclusion to be reached. Either way, *K*-trap rotations are providing a crucial testing ground for nuclear models, and for understanding the way that small fermion systems rotate.

The decay of spin traps

The key to knowing where and how to find the most highly excited spin traps is to understand what controls their decay rates. Some aspects^{37,38} are straightforward: higher-multipolarity electromagnetic transitions are slower, and lower-energy transitions are slower (Box 1).

It is also clear that, for *K*-traps, the quantum number *K* plays

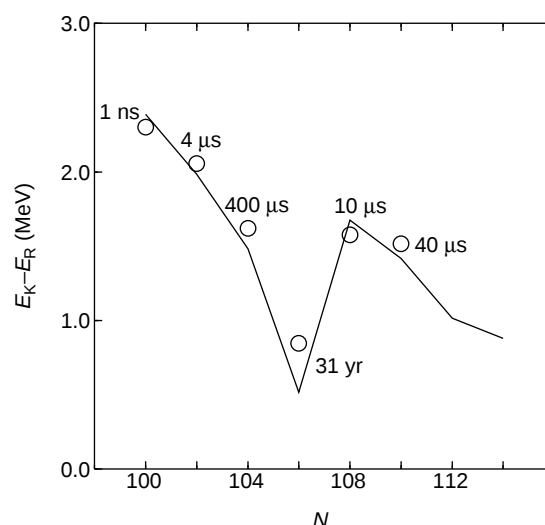


Figure 6 Excitation energy of even-mass hafnium isomers relative to a rotor as a function of neutron number. The circles, with half-lives indicated, represent the lowest-energy experimentally observed four-quasiparticle isomers. The solid line indicates the corresponding calculated values, which extend to higher neutron numbers than can currently be accessed experimentally. The calculations are based on the method of Jain *et al.* (ref. 49), with fixed pairing forces.

an important role. If the transition multipolarity is subtracted from the change in K resulting from the transition, then the degree of forbiddenness, $\nu = \Delta K - \lambda$, is obtained. Löbner⁹ found that, empirically, for each extra unit of forbiddenness, the transition rate reduces by a factor of about 100, which may be expressed as $f_\nu = F_W^{1/\nu} \approx 100$ (here F_W is the ratio of the Weisskopf estimate—see Box 1—to the measured transition rate). The term f_ν is called the ‘reduced hindrance’, and its value depends on the nuclear structure.

Over the past decade, some notable examples have been found^{39–45} where the reduced hindrance appears to be very small ($2 < f_\nu < 10$), calling into question the validity of the K quantum number, and leaving open the question of how to understand K -trap half-lives. One of the earliest cases^{39,40} of very small f_ν , and still the most striking⁴³, was seen in the decay of a 3.3-MeV isomer in ¹⁷⁹W. The isomer appeared to decay predominantly with a K change of 14 units and $f_\nu = 2$. Compared to the rule of thumb of $f_\nu = 100$, the transition goes at a rate that is a factor of 10^{20} too fast.

One explanation^{41,44,46} for such rapid transitions came in the form of a new mode of nuclear decay: tunnelling through axially asymmetric shapes. Instead of a nuclear reorientation being needed to accommodate the change of K value, a shape change could achieve the same ends. Even though the initial and final shapes look much the same (see, for example, Fig. 1c), the tunnelling transition involves passing through radically different intermediate shapes. Tunnelling calculations have proved remarkably successful in reproducing observed transition rates⁴⁷.

A different viewpoint^{42,43,45} invokes two different kinds of K mixing. One is a Coriolis effect that perturbs the states to which an isomer decays (such states typically involve rapid collective rotation, as mentioned earlier). The other effect depends on the excitation energy of an isomer. It has been demonstrated⁴⁵ that, for certain isomer decays, there is a systematic decline in reduced hindrance as the difference increases between the isomer excitation energy, and the expected energy of a typical rotational state of the same spin. The correlation (Fig. 5) can be explained⁴⁵ by statistical K mixing that depends on the density of states. States with the same spin as an isomer, but having lower K values, are assumed to mix and introduce low- K components into the isomeric state. The rapid increase in the number of such states per unit energy interval, as the excitation energy of the isomer increases, can be used to estimate the energy dependence of the declining reduced hindrance. This is seen (the curve in Fig. 5) to follow the trend of the empirical values. The lowest reduced-hindrance values, which are seen to fall significantly below the density-of-states estimate, can be explained by the additional influence of the Coriolis effect, as mentioned above.

The outcome is that the internal energy of an isomer, measured relative to the energy of a simple rotor, is the key feature in determining the K -purity. With regard to half-life, the excitation energy is therefore doubly important: first, because lower-energy transitions are slower, and second because lower-energy isomers have less K mixing—so the transitions are slower again. Excitation energies that are high in an absolute sense on a nuclear scale, yet low relative to a rotor, allow high-spin isomers to have both long half-lives and high internal energies. This makes them attractive as a potential means of energy storage (see the section ‘Quest for stimulated energy release’).

Spin-trap predictions

The reliability of predicting where multi-quasiparticle spin traps may be found—in which nuclei and at what energies—has improved considerably in recent years. Early work by Aberg¹⁷ provided a valuable guide, but pairing effects were treated only qualitatively. More recently, a methodology has been developed^{18,48,49} for treating BCS pairing at fixed shapes, with residual nucleon–nucleon interactions included in a semi-empirical manner. This, together with more realistic pairing treatments³⁴ and allowance for configuration (nucleon orbit) dependent shapes within a given nucleus⁵⁰, has

enabled quantitative estimates of spin-trap energies to be made, although with significant uncertainties (up to about 200 keV).

Half-life predictions are more problematic, because it is not yet possible to calculate realistically the amplitude of the low- K and high- K extremes of K -mixed wavefunctions. Shape-tunnelling estimates treat only one mode of decay and therefore give a limit⁴⁷, but the uncertainties in excitation energies can translate into uncertainties in level-orderings, which can profoundly affect decay rates.

Nevertheless, it is instructive to calculate the excitation energies of a series of isomers, relative to a rotor. This is shown in Fig. 6 for four-quasiparticle isomers of various hafnium isotopes: the lowest-energy data point corresponds to the isomer in ¹⁷⁸Hf₁₀₆ with a 31-year half-life. The other measured half-lives that are given in Fig. 6 illustrate that the relatively higher-energy isomers are shorter lived, with many orders of magnitude decrease in half-life for a relative energy increase of a few hundred keV. For the heavier isotopes, low excitation energies relative to rotor values are predicted (see solid curve), so that the discovery of more extremely long-lived isomers may be anticipated. This sets a clear challenge for future experiments. Indeed, the data points for ¹⁸⁰Hf₁₀₈ and ¹⁸²Hf₁₁₀ in Fig. 6 are new results²¹ obtained by bombarding ¹⁸⁰Hf nuclei with a beam of 1.6-GeV ²³⁸U. The same technique is capable of probing further into the neutron-rich hafnium isotopes and other adjacent elements¹⁰ ($Z = 70$ –76), where many long-lived isomers are expected.

Such calculations not only reveal what sort of energies may be expected in four-quasiparticle states, but also predict higher-excitation-energy, higher-spin isomers with more quasiparticles, many of them in nuclei with stable (non-radioactive) ground states. For example, there is expected to be an eight-quasiparticle $K = 30$ isomer in ¹⁷⁸Hf at about 7 MeV, which could be made by bombarding ⁴⁸Ca with a beam of ¹³⁴Te at an energy that causes four neutrons to be evaporated. But ¹³⁴Te is radioactive, with a half-life of 42 minutes. This experiment must await the completion of radioactive-beam accelerator facilities¹³ that are already planned or under construction, when such exotic beams might become available.

The quest for stimulated energy release

Nuclear isomers provide a form of energy storage. Whereas β -decaying radioactive nuclei can also store MeV energies (per nucleus) for long periods of time, nuclear isomers usually decay electromagnetically, by γ -ray emission or internal conversion. This difference may offer improved opportunities for stimulating the decay of isomers.

The interaction of visible or ultraviolet photons with nuclei has long been sought⁵¹. In principle, low-energy ($\sim$ eV) photons should be able to initiate isomer decay, either by stimulated emission to a nearby lower-energy state, or by stimulated absorption to a nearby higher-energy state. In either case, if the stimulated transition is to a short-lived energy level, a cascade of γ -rays could be released, with a total energy of several MeV. If this scheme were to be realized, then it would be possible to have a type of nuclear reservoir, where the energy could be released with a photon ‘switch’. It would not be necessary to wait for the intrinsic half-life to release the energy and, if the isomer were in a nucleus with a stable ground state, there would be no subsequent radioactive waste. Moreover, the potential development of a γ -ray laser would be brought a significant step closer⁵². Such an idealized scenario has not yet been realized, and low-energy photons have yet to be proven to stimulate nuclear transitions; but this does not mean that there are no possibilities.

In the search for stimulated transitions, the isomers studied to date have mostly been at low energies (that is, low on a nuclear energy scale). The 45-hour isomer in ²²⁹Th, at an excitation energy of 3.5 eV, has received special attention^{51,53,54} because the nuclear excitation is on a similar energy scale to atomic valence electrons. Although the stimulation of the 3.5-eV transition would release no additional energy, it is of fundamental interest for investigating photon–nucleus interactions. But the MeV isomers discussed here

have the potential to release significant quantities of energy. A recent encouraging advance was the de-excitation⁵⁵, using <100-keV photons, of the 2.4-MeV, 31-year isomer in ¹⁷⁸Hf, although this is still a long way from de-excitation with eV photons.

At face value, the use of eV photons to stimulate nuclear transitions requires one or more states to lie within a few eV of the isomer. For low-lying isomers, this presents a general problem. A chance near-degeneracy is needed, such as that found exceptionally in ²²⁹Th. For highly excited isomers, however, with energies of several MeV, the isomers become embedded in a high statistical density of excited states. For example, at 5 MeV in a nucleus such as ¹⁷⁸Hf, there is on average about one state per eV, so the 'chance' of a near-degeneracy becomes certain.

Notwithstanding this certainty, the isomer would probably have very different spin quantum numbers from the background of statistical states (otherwise it would not be an isomer in the first place). To stimulate isomer de-excitation, additional processes may therefore be required; one such example is the assisted electron-bridge mechanism⁵¹, in which excitation of atomic electrons can eliminate energy and spin mismatching between the isomeric state of a nucleus and the state to which it decays.

There is, therefore, room for chance to play an important role in providing the right conditions for isomer formation, and for allowing transitions to be stimulated by low-energy photons. Many new isomers are predicted to exist in nuclei with about 180 nucleons. The insights already gained into the nucleus, together with the development of experimental facilities, suggest the possibility of wide technological benefits in the future. □

Philip Walker is at the Department of Physics, University of Surrey, Guildford GU2 5XH, UK. George Dracoulis is at the Department of Nuclear Physics, Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 0200, Australia.

1. Hansen, P. G., Jensen, A. S. & Jonson, B. Nuclear halos. *Annu. Rev. Nucl. Part. Sci.* **45**, 591–634 (1995).
2. Nolan, P. J. & Twin, P. J. Superdeformed shapes at high angular momentum. *Annu. Rev. Nucl. Part. Sci.* **38**, 533–562 (1988).
3. Hofmann, S. New elements—approaching $Z = 114$. *Rep. Prog. Phys.* **61**, 639–689 (1998).
4. Audi, G., Bersillon, O., Blachot, J. & Wapstra, A. H. The Nubase evaluation of nuclear and decay properties. *Nucl. Phys. A* **624**, 1–124 (1997).
5. Meitner, L. & Frisch, O. R. Disintegration of uranium by neutrons: a new type of nuclear reaction. *Nature* **143**, 239–240 (1939).
6. Britt, H. C. Properties of fission isomers. *At. Data Nucl. Data Tables* **12**, 407–417 (1973).
7. de Voigt, M. J. A., Dudek, J. & Szymanski, Z. High-spin phenomena in atomic nuclei. *Rev. Mod. Phys.* **55**, 949–1046 (1983).
8. Byrne, A. P. *et al.* 34 μ s isomer at high spin in ²¹²Fr: evidence for a many-particle octupole coupled state. *Phys. Rev. C* **42**, R6–R9 (1990).
9. Lobner, K. E. G. Systematics of absolute transition probabilities of K-forbidden gamma-ray transitions. *Phys. Lett. B* **26**, 369–370 (1968).
10. Wheldon, C. *et al.* Opening up the $A \approx 180$ K-isomer landscape: inelastic excitation of new multi-quasiparticle yrast traps. *Phys. Lett. B* **425**, 239–245 (1998).
11. Kondev, F. G. *et al.* Configuration changes and hindered decays in four- and six-quasiparticle isomers in ¹⁷⁸Ta. *Phys. Rev. C* **54**, R459–R463 (1996).
12. Gjorup, N. L. *et al.* Seven and nine quasi-particle K-isomers in ¹⁷⁵Hf. *Zeit. Phys. A* **337**, 353–354 (1990).
13. Proceedings of the International Workshop on Physics with Radioactive Beams. (eds Sinha, B., Shyam, R. & Thompson, I. J.) *J. Phys. G* **24**, 1309–1656 (1998).
14. Khoo, T. L. & Lovhoiden, G. Structural changes in the yrast states of ¹⁷⁸Hf. *Phys. Lett. B* **67**, 271–274 (1977).
15. Boos, N. *et al.* Nuclear properties of the exotic high-spin isomer ¹⁷⁸Hf_{m2} from collinear laser spectroscopy. *Phys. Rev. Lett.* **72**, 2689–2692 (1994).

16. Deylitz, S. *et al.* Inelastic deuteron scattering from the high-spin isomer ¹⁷⁸Hf_{m2} (16⁺). *Phys. Rev. C* **53**, 1266–1272 (1996).
17. Aberg, S. An investigation of yrast traps in some prolate Hf isotopes. *Nucl. Phys. A* **306**, 89–100 (1978).
18. Jain, K., Walker, P. M. & Rowley, N. Residual interactions in multi-quasiparticle states. *Phys. Lett. B* **322**, 27–32 (1994).
19. Dracoulis, G. D. Trends in nuclear structure with heavy ions. *Nucl. Instrum. Methods Phys. Res. A* **382**, 1–19 (1996).
20. Mullins, S. M. *et al.* Rotational band on the 31 yr 16⁺ isomer in ¹⁷⁸Hf. *Phys. Lett. B* **393**, 279–284 (1997).
21. D'Alarcao, R. *et al.* High-K isomers in neutron-rich hafnium nuclei at and beyond the stability line. *Phys. Rev. C* **59**, R1227–R1231 (1999).
22. Cocks, J. F. C. *et al.* Observation of octupole structures in radon and radium isotopes and their contrasting behavior at high spin. *Phys. Rev. Lett.* **78**, 2920–2923 (1997).
23. Dasso, C. H., Pollaro, G. & Winther, A. Systematics of isotope production with radioactive beams. *Phys. Rev. Lett.* **73**, 1907–1910 (1994).
24. Grzywacz, R. *et al.* Identification of μ s-isomers produced in the quasifragmentation of a ¹¹²Sn beam. *Phys. Lett. B* **355**, 439–446 (1995).
25. Pfutzner, M. *et al.* New isotopes and isomers produced by the fragmentation of ²³⁸U at 1000 MeV/nucleon. *Phys. Lett. B* **444**, 32–37 (1998).
26. Bohr, A. & Mottelson, B. R. Rotational states in even-even nuclei. *Phys. Rev.* **90**, 717–719 (1953).
27. Burson, S. B., Blair, K. W., Keller, H. B. & Wexler, S. The radiations from hafnium. *Phys. Rev.* **83**, 62–68 (1951).
28. Bohr, A. & Mottelson, B. R. *Nuclear Structure* Vol. II (Benjamin, New York, 1975).
29. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
30. Belyaev, S. T. Effect of pairing correlations on nuclear properties. *Mat. Fys. Medd. Dan. Vid. Selsk.* **31**(11), 1–55 (1959).
31. Shimizu, Y. R., Garrett, J. D., Broglia, R. A., Gallardo, M. & Vigezzi, E. Pairing fluctuations in rapidly rotating nuclei. *Rev. Mod. Phys.* **61**, 131–168 (1989).
32. Purry, C. S. *et al.* Rotation of an eight-quasiparticle isomer. *Phys. Rev. Lett.* **75**, 406–409 (1995).
33. Dracoulis, G. D., Kondev, F. G. & Walker, P. M. Pairing reduction and rotational motion in multi-quasiparticle states. *Phys. Lett. B* **419**, 7–13 (1998).
34. Burglin, O. & Rowley, N. Exact pairing calculations in a large Nilsson-model basis: comparison with BCS and Lipkin-Nogami methods. *Nucl. Phys. A* **602**, 21–40 (1996).
35. Pashkevich, V. V. & Frauendorf, S. Effect of the shell structure on the moment of inertia; behavior of the averaged moment of inertia. *Sov. J. Nucl. Phys.* **20**, 588–592 (1975).
36. Frauendorf, S. in *Proc. Workshop on Gammasphere Physics* 272–283 (World Scientific, Singapore, 1995).
37. von Weizsacker, C. F. Metastable Zustände der Atomkerne. *Naturwissenschaften* **24**, 813–814 (1936).
38. Blatt, J. M. & Weisskopf, V. F. *Theoretical Nuclear Physics* (Wiley & Sons, New York, 1952).
39. Bernthal, F. *et al.* Connection between backbending and high-spin isomers decay in ¹⁷⁹W. *Phys. Lett. B* **74**, 211–214 (1978).
40. Pedersen, J., Bjornholm, S., Borggreen, J., Kownacki, J. & Sletten, G. Intrinsic excitations as yrast structures in ¹⁷⁹W at $I \sim 35/2\hbar$. *Physica Scripta T* **5**, 162–164 (1983).
41. Chowdhury, P. *et al.* Decay of high-spin isomers in Os nuclei by barrier penetration. *Nucl. Phys. A* **485**, 136–160 (1988).
42. Walker, P. M. *et al.* High-K barrier penetration in ¹⁷⁴Hf: a challenge to K selection. *Phys. Rev. Lett.* **65**, 416–419 (1990).
43. Walker, P. M. *et al.* Resolution of the ¹⁷⁹W isomer anomaly: exposure of a Fermi-aligned s band. *Phys. Rev. Lett.* **67**, 433–436 (1991).
44. Crowell, B. *et al.* Novel decay modes of high-K isomers: tunneling in a triaxial landscape. *Phys. Rev. Lett.* **72**, 1164–1167 (1994).
45. Walker, P. M. *et al.* K-forbidden transitions from multi-quasiparticle states. *Phys. Lett. B* **408**, 42–46 (1997).
46. Bengtsson, T. *et al.* Competition between pairing-assisted tunneling and single-particle alignment in the decay of high-K isomeric states. *Phys. Rev. Lett.* **62**, 2448–2451 (1989).
47. Narimatsu, K., Shimizu, R. & Shizuma, T. γ -tunneling calculations for the decays of the K isomers in the Hf, W, Os region. *Nucl. Phys. A* **601**, 69–88 (1996).
48. Walker, P. M. *et al.* Multi-quasiparticle and rotational structures in ¹⁷⁹W: Fermi alignment, the K-selection rule and blocking. *Nucl. Phys. A* **568**, 397–444 (1994).
49. Jain, K. *et al.* Multi-quasiparticle states in the mass-180 region. *Nucl. Phys. A* **591**, 61–84 (1995).
50. Xu, F. R., Walker, P. M., Sheikh, J. A. & Wyss, R. Multi-quasiparticle potential-energy surfaces. *Phys. Lett. B* **435**, 257–263 (1998).
51. Matinyan, S. Lasers as a bridge between atomic and nuclear physics. *Phys. Rep.* **298**, 199–249 (1998).
52. Baldwin, G. C. & Solem, J. C. Recoilless gamma-ray lasers. *Rev. Mod. Phys.* **69**, 1085–1117 (1997).
53. Utter, S. B. *et al.* Reexamination of the optical gamma ray decay in ²²⁹Th. *Phys. Rev. Lett.* **82**, 505–508 (1999).
54. Shaw, R. W., Young, J. P., Cooper, S. P. & Webb, O. F. Spontaneous ultraviolet emission from ²³³Uranium/²²⁹Thorium samples. *Phys. Rev. Lett.* **82**, 1109–1111 (1999).
55. Collins, C. B. *et al.* Accelerated emission of gamma rays from the 31-year isomer of ¹⁷⁸Hf induced by x-ray irradiation. *Phys. Rev. Lett.* **82**, 695–698 (1999).
56. Rowe, D. J. *Nuclear Collective Motion: Models and Theory* (Methuen, London, 1970).

Correspondence and requests for materials should be addressed to P.W. (e-mail: p.walker@surrey.ac.uk).

Possible long-lived asteroid belts in the inner Solar System

N. Wyn Evans & Serge Tabachnik

Theoretical Physics, Department of Physics, 1 Keble Road, Oxford OX1 3NP, UK

Recent years have seen the discovery of several objects in stable orbits in the outer Solar System^{1–3}; these bodies include objects in the Kuiper belt (also known as the Kuiper–Edgeworth belt) as well as the Centaurs. Moreover, another region of orbital stability has been identified between the orbits of Uranus and Neptune⁴. Here we report evidence from numerical simulations of zones of orbital stability in the inner Solar System. We find that there are two possible long-lived belts of asteroids. The first region lies between the Sun and Mercury, in the range 0.09–0.21 astronomical units, where remnant planetesimals may survive for the age of the Solar System provided that their radii are greater than ~0.1 kilometres. The second region of stability is between Earth and Mars (range 1.08–1.28 astronomical units), where a population of bodies that are on circular orbits may survive. A search through the catalogues of near-Earth objects reveals an excess of asteroids with low eccentricities and inclinations occupying this latter region: several examples are the recently discovered objects 1996 XB27, 1998 HG49 and 1998 KG3.

Symplectic integrators^{5,6} with individual timesteps⁷ provide fast algorithms that are perfectly suited to long numerical integrations

of low-eccentricity orbits in a nearly keplerian force field. Individual timesteps are a great help for our work, as orbital clocks ‘tick’ much faster in the inner Solar System than in the outer. In the simulations that we report here, more than 1,000 test particles are distributed on concentric rings with values of the semimajor axis between 0.1 and 2.2 astronomical units (AU). Each of these rings is located in the invariable plane and hosts five test particles with starting longitudes $n \times 72^\circ$ with $n = 0, \dots, 4$. Initially, the inclinations and eccentricities vanish for the whole sample of test particles. The test particles are perturbed by the Sun and planets but do not themselves exert any gravitational forces. The full gravitational effects of all the planets (except Pluto) are included. The initial positions and velocities of the planets, as well as their masses, come from the JPL Planetary and Lunar Ephemerides DE405. For all the computations, the timestep for Mercury is 14.27 d. The timesteps of the planets are in the ratio 1:2:2:4:8:8:64:64 for Mercury moving outwards through to Neptune. The relative energy error is oscillatory and has a peak amplitude of $\sim 10^{-6}$ over the 100-Myr integration timespans (compare refs 6, 8). After each timestep, the test particles are examined. If their orbits have become hyperbolic or have entered the sphere of influence of any planet⁹ or have approached closer than ten solar radii to the Sun, they are removed from the simulation. This general procedure is familiar from a number of recent studies on the stability of test particles in the Solar System^{4,10}. For example, Holman⁴ uncovered evidence for a possible belt between Uranus and Neptune by a similar integration of test particles in the gravitational field of the Sun and the four giant planets. His integrations reached the impressive timescale of 4.5 Gyr—of the order of the age of the Solar System. Simulations of the inner Solar System are much more laborious, as the orbital

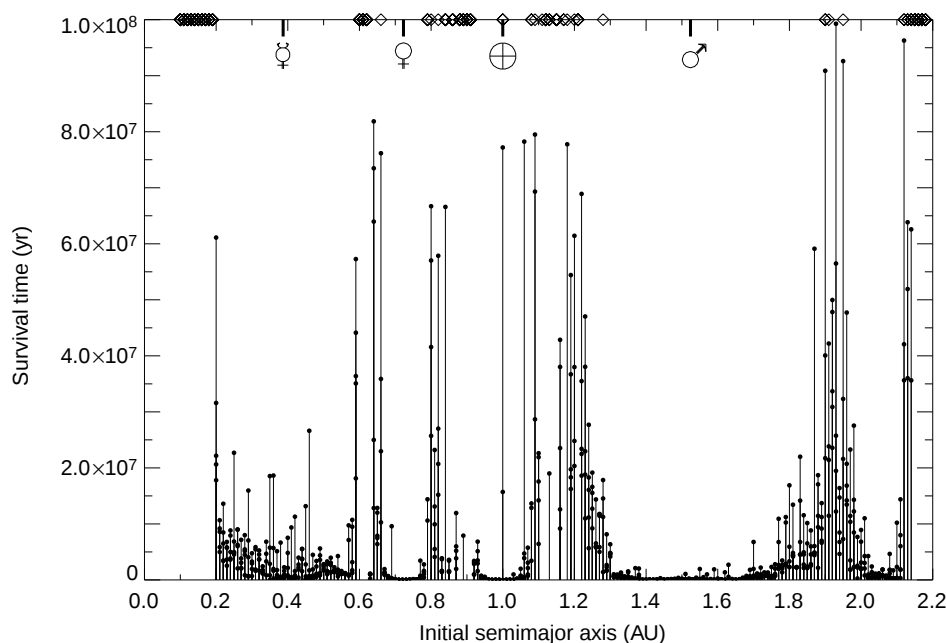


Figure 1 The calculated survival time plotted against the starting semimajor axis (in astronomical units) for test particles in the inner Solar System. At each semimajor axis, five test particles are launched at equally spaced longitudes and their initially circular orbits are integrated for 100 Myr. The times of ejection are marked with filled circles and joined with solid vertical lines. Test particles close to the terrestrial planets are rapidly removed. However, between all the terrestrial planets, there are narrow belts of stable circular orbits that survive for the full duration of integration of 100 Myr. The semimajor axes of the surviving test particles are marked by open diamonds at the top of the figure, together with symbols marking the locations of the four terrestrial planets. (Left to right,

Mercury, Venus, Earth, Mars.) Test particles survive in the following four regions—between the Sun and Mercury (0.1–0.19 AU), between Mercury and Venus (0.6–0.66 AU), between Venus and the Earth (0.79–0.91 AU) and between the Earth and Mars (1.08–1.28 AU). The two test particles that remain after 100 Myr at an initial semimajor axis of 1 AU are actually librating about the Earth’s Lagrange points. (These calculations were performed on personal computers that employ 80 bits internally and which offer the option of compilation in long double precision with a 64 bit mantissa. At least 64 bits are required to keep Mercury’s longitude error below 0.01 rad over 100-Myr timescales (see ref. 21). Standard double precision offers a mantissa of just 53 bits.)

Table 1 Simulation results

Belt	<i>a</i>	<i>b</i>	<i>N</i> ₀	<i>N</i> _{exp} (1 Gyr)	<i>N</i> _{exp} (5 Gyr)
V	472.34 ± 5.0	− 26.03 ± 0.68	300	238	220
M–V	681.7 ± 3.7	− 69.73 ± 0.52	250	54	5
V–E	1041.1 ± 3.9	− 111.51 ± 0.56	375	38	–
E–M	1229.2 ± 3.2	− 118.26 ± 0.46	500	165	82
Belt	<i>c</i>	<i>d</i>	<i>N</i> ₀	<i>N</i> _{exp} (1 Gyr)	<i>N</i> _{exp} (5 Gyr)
V	2.743 ± 0.008	0.040 ± 0.001	300	241	226
M–V	3.445 ± 0.017	0.167 ± 0.002	250	88	67
V–E	3.725 ± 0.007	0.189 ± 0.001	375	106	78
E–M	3.531 ± 0.008	0.133 ± 0.001	500	215	173

Fitting parameters *a*, *b*, *c*, *d* are defined in equation (1).

The belt labels V, M–V, V–E and E–M refer to the Vulcanoids (0.09–0.21 AU), the Mercury–Venus belt (0.58–0.68 AU), the Venus–Earth belt (0.78–0.93 AU) and the Earth–Mars belt (1.08–1.28 AU), respectively. Test particles are placed at semimajor axes separated by 0.002 AU in each of the belts and the orbits are re-simulated for 100 Myr. The initial number of test particles in each belt, *N*₀, is given. In each case, the data *N*(*t*) is fitted for 1 Myr < *t* < 100 Myr. The parameters in the logarithmic and the power-law fits to the number of remaining test particles *N*(*t*) after time *t* are listed in the upper and lower tables. The last two columns of the tables give the extrapolated number of test particles estimated to remain after 1 and 5 Gyr. The estimated uncertainties in the fitted parameters indicate that the logarithmic law is a better guide for extrapolation—it is also more pessimistic than the power-law fit. This suggests that only the Vulcanoid orbits and the Earth–Mars belt can be accepted as candidate repositories for long-lived objects. (The simulation of the Vulcanoid belt has been performed in long double precision on personal computers, the remaining three belts in double precision on a supercomputer.)

period of Mercury is ~88 d (as compared to ~4,332 d for Jupiter, the giant planet with the shortest orbital period). This forces us to use a much smaller timestep and round-off error becomes an obstacle to believable results. We therefore adopt the strategy of running the simulations on nearly 20 personal computers of varying processor speeds, so that the calculations are performed in long double precision implemented in hardware. The integration of the orbits of 1,050 test particles for 100 Myr occupied these computers for over four months.

Figure 1 shows the results of these calculations: the survival times of the test particles are plotted against starting semimajor axis. There are five test particles at each starting position, so the vertical lines in the figure join five filled circles which mark their ejection times. The locations of particles that survive for the entire 100-Myr timespan are marked by diamonds on the upper horizontal axis. Around each of the terrestrial planets, there is a swathe of test particles that are ejected rapidly on a precession timescale. This band is much broader around Mars than around the Earth or Venus, perhaps because of the higher eccentricity of Mars' orbit. There are also narrow belts of test particles that survive for the full integration. So, for example, all 50 of the test particles with starting semimajor axes between 0.1 and 0.19 AU are still present at the end of the 100-Myr integration. The existence of a population of small asteroid-like bodies—known as the Vulcanoids—wandering in intra-mercurial orbits has been hypothesized before^{11–13}. There are 16 surviving test particles with starting semimajor axes between 0.6 and 0.66 AU, suggesting the possible existence of a narrow belt between Mercury and Venus. A somewhat larger third belt of 33 surviving test particles occupies a belt between Venus and the Earth; their starting semimajor axes range from 0.79 to 0.91 AU. Finally, there is a broad belt between the Earth and Mars from 1.08 to 1.28 AU in which a further 26 test particles survive. The possibility of the existence of belts between Venus and the Earth and between the Earth and Mars was raised by Mikkola and Innanen¹⁰ on the basis of 3-Myr integrations.

Of course, these results must be treated with considerable reserve, as 100 Myr is just ~2% of the age of the Solar System since the assembly of the terrestrial planets (~5 Gyr ago). We estimate that if this simulation in long double precision were to be continued till the integration time reaches even 1 Gyr, then it would consume ~3.5 yr of time. Accordingly, we use the standard, albeit approximate, device of re-simulating with greater resolution and extrapolating the results. At semimajor axes separated by 0.002 AU, five test particles are again launched on initially circular orbits with starting longitudes *n* × 72° with *n* = 0, ..., 4. The number of test particles *N*(*t*) remaining after time *t* is monitored for each of the four belts. Table 1 gives the results of fitting the data between 1 and 100 Myr to

the following logarithmic and power-law decays

$$N(t) = a + b \log_{10}(t), \quad N(t) = \frac{10^c}{t^d} \quad (1)$$

where *t* is in years. In the last two columns of Table 1, the expected number of test particles remaining after 1 and 5 Gyr is computed. The uncertainties in the fitted parameters suggest that the logarithmic fall-off is a better—and more pessimistic—fit to the asymptotic behaviour than a power law (compare refs 4, 14). Our extrapolations suggest that two of the belts—those lying between Mercury and Venus, and between Venus and the Earth—will become almost entirely depleted after 5 Gyr. However, even taking a very pessimistic outlook, it seems certain that some of the Vulcanoids and some of the test particles with starting semimajor axes between 1.08 and 1.28 AU in the Earth–Mars belt may survive for the full age of the Solar System. We can make a crude estimate of present-day numbers by extrapolation from the main belt asteroids (compare ref. 4). Assuming that the primordial surface density falls inversely with distance, we find that the Earth–Mars belt may be occupied by perhaps ~1,000 remnant objects. Of course, these objects are now outnumbered by the more recent arrivals, the asteroids ejected via resonances from the main belt, which may number a few thousand in total¹⁵.

A systematic search for Vulcanoids has already been conducted by Leake *et al.*¹², who exploited the fact that bodies so close to the Sun are identifiable from their substantial infrared excess. No candidate objects were found. However, the survey was limited to a small area of just 6 square degrees and was estimated to be ~75% efficient to detection of bodies brighter than fifth magnitude in the L band. This result places constraints on the existence of a population of objects with radii greater than ~50 km, but minor bodies with the typical sizes of small asteroids (~10 km) will have evaded detection. On theoretical grounds, Vulcanoids have been proposed to resolve apparent contradictions between the geological and the geophysical evidence on the history of the surface features on Mercury^{11,12}. The robustness of the Vulcanoid orbits partly stems from the fact that there is only one neighbouring planet and so may be compared to the stability of the Kuiper–Edgeworth belt. Even after 100 Myr, some 80% of our Vulcanoid orbits still have eccentricities *e* < 0.2 and inclinations *i* < 10°. It is this evidence, together with the low rate of attrition of their numbers, that suggests that they can continue for times of the order of the age of the Solar System. The outer edge of the Vulcanoid belt is at ~0.21 AU. Objects beyond this are dynamically unstable, and are excited into Mercury-crossing orbits on 100-Myr timescales. The inner edge of the belt is not so sharply defined. Small objects close to the Sun may be susceptible to destruction both by Poynting–Robertson drag¹⁶ and by

evaporation¹⁷. Taking the mean density and Bond albedo of a typical Vulcanoid to be the same as that of Mercury, we find that objects with radii z satisfying $0.1 \leq z \leq 50$ km can evade both drag and evaporation in the Vulcanoid belt. This is one of the most dynamically stable regimes in the entire Solar System. If further searches do not detect any intra-mercurial objects, this is a strong indicator that other processes—such as planetary migrations—may have disrupted the population.

Although there are no known intra-mercurial bodies, we can find candidate objects for the Earth–Mars belt. Suppose we search an asteroidal database¹⁸ for objects with inclinations $i < 10^\circ$ and eccentricities $e < 0.2$ between the semimajor axes of Earth and Mars, then we find that there are ten objects. Of these, seven lie within our suggested Earth–Mars belt (1.08–1.28 AU), which is evidence for an enhancement of nearly circular orbits in this region. An even more striking test is to search through the objects between the Earth and Mars for asteroids of low eccentricity and inclination that are not planet-crossing. There are only three such objects (1996 XB27, 1998 HG49 and 1998 KG3) among all the asteroids in this database, and all three lie between 1.08 and 1.28 AU. Most of the ~ 50 asteroids with semimajor axes at present located in our Earth–Mars belt are moving on orbits with large eccentricities and inclinations. They are not dynamically stable, and will evolve on timescales of the order of a few million years. Most of these objects are believed to be asteroids ejected from resonance locations in the main belt, although a handful may even be comets whose surfaces have become denuded of volatiles¹⁹. However, the seeming enhancement of circular orbits in this region hints at a primordial population whose orbits are very slightly eccentric and slightly inclined. Ejection from the main belt will tend to increase the eccentricity of an asteroid²⁰. So the slightly eccentric objects may well be remnant planetesimals, the original denizens of the region before it was colonized by asteroids from the resonance locations in the main belt. $\square$

Received 26 October 1998; accepted 12 March 1999.

- Kowal, C. T. in *Asteroids* (ed. Gehrels, T.) 436–439 (Univ. Arizona Press, Tucson, 1979).
- Jewitt, D. & Luu, J. X. Discovery of the candidate Kuiper belt object 1992 QB1. *Nature* **362**, 730–732 (1993).
- Williams, I. P., O'Cellaigh, D. P., Fitzsimmons, A. & Marsden, B. G. The slow-moving objects 1993 SB and 1993 SC. *Icarus* **116**, 180–185 (1995).
- Holman, M. J. A possible long-lived belt of objects between Uranus and Neptune. *Nature* **387**, 785–788 (1997).
- Wisdom, J. & Holman, M. J. Symplectic maps for the n-body problem. *Astron. J.* **102**, 1528–1538 (1991).
- Saha, P. & Tremaine, S. D. Symplectic integrators for solar system dynamics. *Astron. J.* **104**, 1633–1640 (1992).
- Saha, P. & Tremaine, S. D. Long term planetary integration with individual time steps. *Astron. J.* **108**, 1962–1969 (1994).
- Holman, M. J. & Wisdom, J. Dynamical stability in the outer solar system and the delivery of short period comets. *Astron. J.* **105**, 1987–1999 (1993).
- Danby, J. M. A. *Fundamentals of Celestial Mechanics* (Willmann-Bell, Richmond, 1988).
- Mikkola, S. & Innanen, K. Solar system chaos and the distribution of asteroid orbits. *Mon. Not. R. Astron. Soc.* **277**, 497–501 (1995).
- Weidenschilling, S. J. Iron/silicate fractionation and the origin of Mercury. *Icarus* **35**, 99–111 (1978).
- Leake, M. A., Chapman, C. R., Weidenschilling, S. J., Davis, D. R. & Greenberg, R. The chronology of Mercury's geological and geophysical evolution—The Vulcanoid hypothesis. *Icarus* **71**, 350–375 (1987).
- Campins, H., Davis, D. R., Weidenschilling, S. J. & Magee, M. in *Completing the Inventory of the Solar System* (eds Rettig, T. W. & Hahn, J. M.) 85–96 (ASP Conf. Proc., Astron. Soc. Pacif., San Francisco, 1996).
- Dones, L., Levison, H. F. & Duncan, M. in *Completing the Inventory of the Solar System* (eds Rettig, T. W. & Hahn, J. M.) 233–244 (ASP Conf. Proc., Astron. Soc. Pacif., San Francisco, 1996).
- Weissman, P. R., A'Hearn, M. F., McFadden, L. A. & Rickman, H. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 880–920 (Univ. Arizona Press, Tucson, 1989).
- Robertson, H. P. Dynamical effects of radiation in the solar system. *Mon. Not. R. Astron. Soc.* **97**, 423–438 (1937).
- Pettit, E. in *Planets and Satellites* (eds Kuiper, G. P. & Middlehurst, B.) 400–427 (Univ. Chicago Press, 1961).
- Bowell, E. G. *The Asteroid Orbital Element Database* at (<http://ftp.lowell.edu/pub/elgb/astorb.html>).
- McFadden, L. A., Tholen, D. J. & Veeder, G. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 442–467 (Univ. Arizona Press, Tucson, 1989).
- Gladman, B. et al. Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277**, 197–201 (1997).
- Saha, P., Stadel, J. & Tremaine, S. D. A parallel integration method for solar system dynamics. *Astron. J.* **114**, 409–414 (1997).

Acknowledgements. We thank P. Saha and S. Tremaine for suggestions and comments, as well as their advice on computational matters, and J. Chambers, M. Holman and L. Dones for helpful criticism. The Royal Society supported the purchase of dedicated computers, and we also thank the Oxford Supercomputing Centre (OSCAR) for support.

Correspondence and requests for materials should be addressed to N.W.E. (e-mail: nwe@thphys.ox.ac.uk).

Observation of mesoscopic vortex physics using micromechanical oscillators

C. A. Bolle*, V. Aksyuk*, F. Pardo*, P. L. Gammel*, E. Zeldov†, E. Bucher*, R. Boie*, D. J. Bishop* & D. R. Nelson†

* Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974, USA

† Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

It has long been known that magnetic fields penetrate type II superconductors in the form of quantized superconducting vortices. Most recent research in this area has, however, focused on the collective properties of large numbers of strongly interacting vortices^{1,2}: the study of vortex physics on the mesoscopic scale (a regime in which a small number of vortices are confined in a small volume) has in general been hampered by the lack of suitable experimental probes. Here we use a silicon micromachined mechanical resonator to resolve the dynamics of single vortices in micrometre-sized samples of the superconductor 2H-NbSe₂. Measurements at and slightly above the lower critical field, H_{c1} (the field at which magnetic flux first penetrates the superconductor), where only a few vortices are present, reveal a rich spectrum of sharp, irreversible vortex rearrangements. At higher fields, where tens of vortices are present, the sharp features become reversible, suggesting that we are resolving a new regime of vortex dynamics in which the detailed configuration of pinning sites, sample geometry and vortex interactions produce significant changes in the measurable vortex response. This behaviour can be described within the framework of interacting vortex lines in a '1 + 1'-dimensional random potential—an important (but largely untested) theoretical model for disorder-dominated systems^{11,12}.

High-Q mechanical oscillators have previously been shown to be powerful probes of condensed matter^{3–5}. However, the recent availability of silicon micromechanics has enormously expanded the power of the technique by allowing high Q and also small size. Figure 1 shows a scanning electron micrograph of our micromachine. The mechanical oscillator consists of a plate connected via two serpentine springs to two support structures attached to the substrate. Under the plate are two electrodes on the silicon substrate. The outline of the electrodes is visible on the top plate owing to print-through during the multiple processing steps. In operation, the device is oscillated torsionally about the long axis of the plate. One electrode is used to drive the plate and the other to detect the actual movement. This torsional mode has a resonant frequency of 45 kHz and a Q of 250,000 at low temperatures. The device was built by using a three-layer polysilicon process at the MCNC foundry (<http://www.mcnc.org>).

The inset to Fig. 1 shows an enlargement of the sample used for the studies described here. It is a single crystal of 2H-NbSe₂ with dimensions of $22 \times 49 \times 1.6 \mu\text{m}^3$. The sample is from a batch of very-high-quality single crystals with typical critical current densities of 100 A cm^{-2} and weak pinning, producing well-ordered flux lattices at low fields as observed with magnetic decoration^{6,7}. Other measurements have confirmed the high quality of these crystals⁸. The T_c of our sample was 7.0 K. The sample was glued onto the oscillator with a microgluing technique that allows us to affix submicrometre-sized particles with precision.

A magnetic field was applied parallel to the face of the crystal and

‡ Permanent address: Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

perpendicular to its rotation axis. A sample with a magnetic moment $\mathbf{M}$ in a field $\mathbf{H}$ exerts an additional torque $\boldsymbol{\tau} = \mathbf{M} \times \mathbf{H}$ on the oscillator, shifting its resonant frequency. The small size of the oscillator allows us to obtain high resonant frequencies despite the soft spring constants, typically $\sim 0.5 \times 10^{-9}$ N m.

Figure 2 shows the frequency and amplitude dependence for a standard zero-field-cooled (ZFC)/field-cooled (FC) temperature sweep. We can see that these oscillators have very little temperature-dependent background, as shown by the black lines. After applying a field of 60 Oe at 5 K we observe a frequency shift of roughly 20 Hz and no change in the oscillator's amplitude. This is a clear signature of the Meissner state. The energy cost of screening a magnetic field applied parallel to the face of a thin superconductor is smaller than the cost of screening the same field applied perpendicular to the face of the sample. The resulting variation in the torque with the tilt angle $d\tau/d\theta$ causes a restoring force that increases the oscillator frequency. The result is an increase in the oscillator frequency with no change in the amplitude (no dissipation). As we warm up the sample (red line) we see a slight decrease in the Meissner restoring force as $\lambda(T)$ starts to become comparable to the sample thickness.

At $T \approx 6.1$ K we see a sudden decrease in the oscillator amplitude and a jump in its frequency. These changes are associated with the first flux line penetration and uniquely define the location of the first penetration field H_p for our sample. We can use the approximation $H_{c1} \approx H_p$ for the lower critical field because the demagnetizing factor for this field direction is very small. Note the increase in the 'noise' in the data that, as we show below, is the resolution of single-vortex events in the oscillator. As we approach T_c both the frequency and the amplitude merge into the background. On cooling of the sample (blue line) in the presence of the 60 Oe field, the flux lines nucleate in the superconductor below T_c , resulting in a frequency shift and a decrease in amplitude (increase in dissipation). Both the frequency shift and the amplitude show a

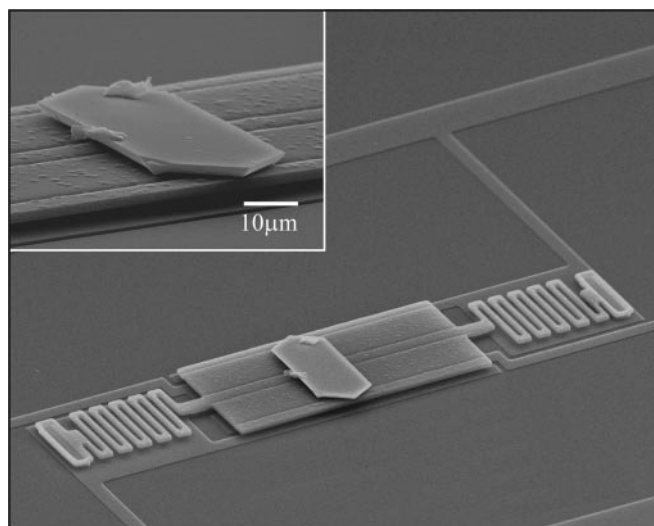


Figure 1 Scanning electron micrograph of a high-Q mechanical oscillator with a hexagonal single crystal of the traditional superconductor 2H-NbSe₂ mounted on top. The oscillator was manufactured in a three-layer polysilicon process. The paddle of the oscillator, formed from the third layer, is freely suspended by two springs anchored to the substrate at its ends. Two electrodes, formed from the first polysilicon layer and located directly under the paddle, are used to excite it in a torsional resonance mode. The gap between the electrodes and the paddle is fixed by the multilayer process and is 2.75 μm . The sample is mounted and glued to the paddle with the help of a micropipette. The inset shows a detail of the sample. The thickness of the sample is slightly larger than the 1.5- μm -thick paddle.

rich structure close to T_c that tends to smooth out at lower temperatures. This structure represents the changes in the pinning strength of the vortex ensemble as the number of flux lines changes.

Figure 3a shows a field sweep close to T_c after a ZFC. The frequency shift in the Meissner state due to the shielding currents up to $H_{c1} \approx 9$ Oe has the expected H^2 dependence, because $\mathbf{M} = (-1/4\pi)\mathbf{H}$ and $\boldsymbol{\tau} = \mathbf{M} \times \mathbf{H}$. Above H_{c1} we see a series of peaks in the frequency that we interpret as the signature of the first several flux lines entering the sample. The field scale and the size of the sample are consistent with this interpretation. As we shall argue below, the first flux lines that go into the sample wander significantly away from the field direction to take advantage of the point disorder. As the field increases, more lines enter the sample and these wandering lines start to interact. Figure 3b shows the response at a lower-temperature ZFC field sweep. The Meissner regime is once again quadratic but the penetration of single vortices above H_{c1} is somewhat different owing to the stronger effects of pinning at lower temperatures. At this temperature, as the individual vortices enter, there are mesas between the jumps. As the field is cycled back and forth a fraction of an oersted, the response in the mesas is reversible but the jumps are not. The reason for the non-monotonic

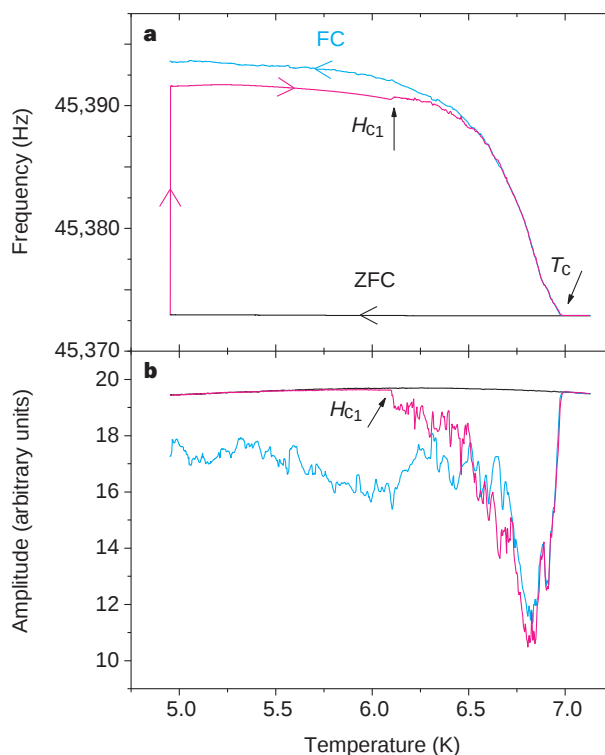


Figure 2 Temperature dependence of the oscillator resonant frequency (**a**) and amplitude (**b**) for a typical zero-field-cooling/field-cooling (FC) experiment. First the system is cooled, in the absence of a magnetic field, to $T = 5$ K and the 'background' frequency and amplitude are recorded, shown in the figure by the black lines. Then a magnetic field $H = 60$ Oe is applied and the temperature is increased to 7.5 K, as shown by the red line. Three different regimes are seen upon heating: (1) the Meissner regime, where shielding currents flow on the surface of the superconductor, yielding an increase in the resonance frequency and no change in amplitude; (2) the lower critical field $H_{c1}(T)$ for flux-line penetration, at which the motion of flux lines inside the superconductor causes dissipation and decreases the oscillation amplitude; (3) the critical temperature, at which both the frequency and the amplitude merge with the background signal. The blue line shows the field-cooling process. Here flux lines form in the superconductor below the critical temperature. The shape of the frequency changes is given by many factors, including the number of flux lines and the pinning strength, giving rise to a complex structure in the oscillator frequency and amplitude.

behaviour with increasing field is the mesoscopic nature of our system. As the vortices enter the sample, sometimes the entire ensemble is better pinned and the frequency increases but sometimes, because of interactions, it is less well pinned and the frequency decreases. The lower trace is $B(H)$ extracted from the data by counting vortices. In this case, for $N(H)$ flux lines in

the sample and a sample cross-section area $S = 22 \times 1.6 \mu\text{m}^2$, we have $B \approx N(H)\Phi_0/S$.

Figure 3c shows a field sweep at a temperature of 6.8 K for both a ramp up (red line) after a ZFC and a ramp down (blue line). Clearly three regimes can be distinguished. Below H_{c1} at ~ 12.5 Oe, the Meissner response is reversible. Above H_{c1} between 12.5 and ~ 30 Oe, the response is irreversible. This is the region where the vortex–vortex interactions are weak enough for the response to depend on field history. Above ~ 30 Oe, one sees a reversible regime in which the response does not depend on field history. In this regime the vortex interactions are strong enough for the ensemble to appear always to find its minimum energy configuration. Note that this regime is still mesoscopic in the sense that the response is highly structured and not monotonic with field. There are ~ 20 flux lines in this part of the curve; the detailed pinning landscape influences how the effective magnetic moment fluctuates with increasing field. This plot is thus a ‘fingerprint’ of the disorder in this specific sample. Taken together, Figs 2 and 3 provide evidence of the mesoscopic (ref. 9) behaviour of a small number of vortices in a finite-sized sample in which the details of the pinning sites produces sample-specific structure in the magnetization of the superconductor.

To describe our results theoretically we determine the constitutive relation $B(H)$ for a low-temperature vortex array interacting strongly with point disorder. We first note that both the in-plane and out-of-plane London penetration depths $\lambda_{\perp}(0) = 69$ nm and $\lambda_c(0) = 230$ nm are comparable to the $d = 1.6 \mu\text{m}$ c -axis sample dimension at the elevated temperatures $T/T_c = 0.95$ of interest here¹⁰. Provided that a vortex feels the confining effect of the surfaces perpendicular to the c axis many times, we can model flux penetration by a collection of confined ‘1 + 1’-dimensional vortex lines stretched out in the field direction y (with dimension $L = 49 \mu\text{m}$) wandering transversely in the x direction (with dimension $W = 22 \mu\text{m}$) and interacting with a random point-like pinning potential. Consider first a single vortex line with trajectory $x(y)$. Adapting the treatment for 2 + 1-dimensional vortices in refs. 11 and 12, we assume that this trajectory occurs with probability proportional to $\exp[-E/T]$ with an energy

$$E = \int_0^L dy \left\{ \frac{1}{2} g^2 \left[\frac{dx(y)}{dy} \right]^2 + V[x(y), y] \right\}$$

where $g = \Phi_0^2/16\pi^2\lambda_{\perp}(T)\lambda_c(T)$ is the vortex tilt modulus in the x direction, Φ_0 the flux quantum and $V(x, y)$ a coarse-grained random pinning potential with disorder average $\overline{V(x, y)V(x', y')} \approx \Delta\delta_{\xi_{\perp}}(x - x')\delta(y - y')$. Here, $\delta_{\xi_{\perp}}(x)$ is a delta function smeared on the scale of the in-plane coherence length $\xi_{\perp}$, and Δ is the effective disorder strength. As pointed out by Huse *et al.*¹³ the associated free energy obeys a noisy Burgers equation that can be solved exactly¹⁴. Although an extensive theoretical literature¹⁵ surrounds this model, our experiment seems to be one of the first to probe this important problem in 1 + 1-dimensional statistical mechanics directly. A key prediction^{13–15} for a vortex trajectory pinned by point disorder reads

$$\langle \Delta x^2(l) \rangle \equiv \langle [x(y+l) - x(y)]^2 \rangle \approx \xi_{\perp}^2 (l/l_c)^{2\xi}$$

where $\xi_{\perp}(0) = 77 \text{ \AA}$, the angular brackets represent a thermal average and $l_{\perp} = \xi_{\perp}(g^2\xi_{\perp}^2/\Delta)$ (we assume $\xi_{\perp} > l_c$). On scales larger than l_c , vortices are strongly pinned with anomalous wandering exponent $\xi = 2/3$. The pinning energy associated with a vortex segment of length $l > l_c$ also grows with l , $U_p(l) \approx g\xi_{\perp}(l/l_c)^{2\xi-1}$. We neglect, throughout, dimensionless constants of order unity. Different results apply for $T > T^*$, where T^* is determined self-consistently from $T^* = (\Delta g(T^*)\xi_{\perp}(T^*))^{1/3}$. For $T^* < T < T_c$, one finds different scale factors, $\langle \Delta x^2(l) \rangle = (Tl_c/g)(l/l_c)^{2\xi}$ and $U_p(l) = T(l/l_c)^{2\xi-1}$ with $l_c(T) = T^3/g\Delta$. For NbSe₂, $(T_c - T^*)/T_c \approx 10^{-5}$, so this regime is not relevant to our experiments.

As more vortices enter the sample, their transverse wandering to optimize the pinning potential is limited by collisions with their

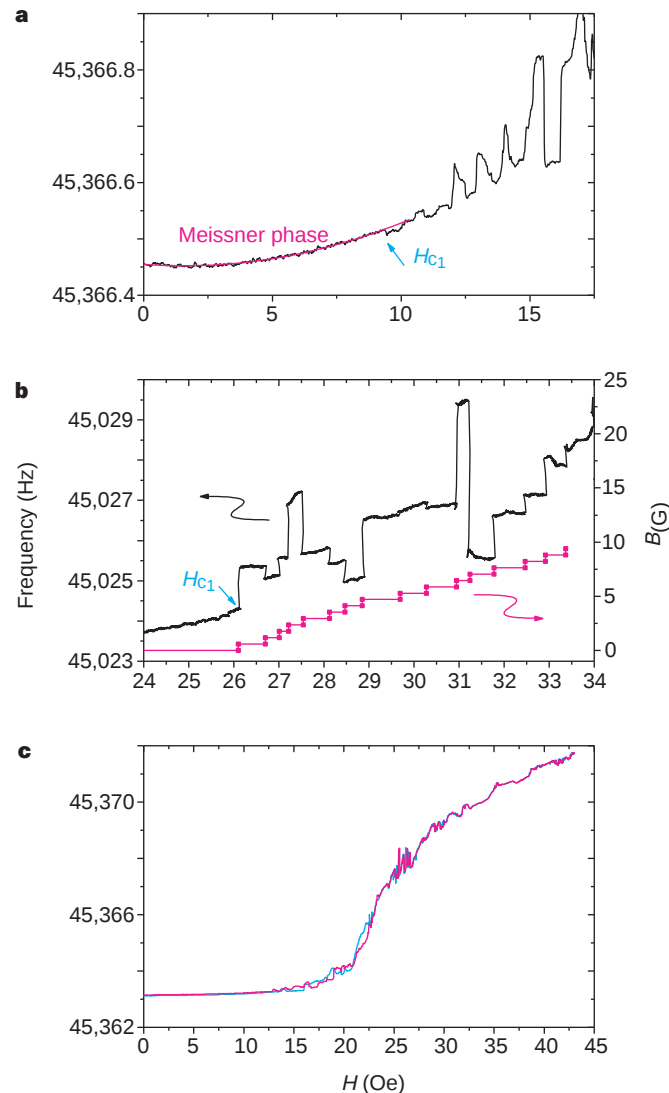


Figure 3 Three different field dependences of the resonant frequency at the fixed temperatures, 6.9 K (**a**), 6.65 K (**b**) and 6.8 K (**c**). **a**, Below $H_{c1} = 9$ Oe we see the H^2 dependence of the frequency shift owing to the Meissner state, highlighted by a polynomial fit in red. Above H_{c1} a series of peaks reflect the penetration of individual flux lines as the field is increased. **b**, the magnetic field was cycled up and down at each point in steps of 0.1 Oe to reveal local irreversibilities. The region below $H = 26$ Oe, corresponding to the Meissner shielding currents, is reversible as expected. Above H_{c1} we obtain reversible mesas separated by irreversible jumps. Each jump is associated with a single flux line entering the sample. The lower trace is $B(H)$ extracted from the data by counting vortices. Note the linear dependence near H_{c1} . **c**, A large-scale field sweep at a fixed temperature. The red line shows the frequency shift as the field is increased, the blue line as it is decreased. Three regions can be distinguished: at low fields, there is a Meissner phase with the expected H^2 change in frequency; above H_{c1} we see an irreversible regime that changes into a reversible one at $H = 30$ Oe. The structure of the resonant frequency above $H = 30$ Oe is given by the number of flux lines, the shape of the sample, the structure of the flux line lattice and the pinning landscape. In this regime the flux line lattice relaxes to exactly the same structure regardless of the previous field history for our field sweep rate of 0.02 Oe s^{-1} .

neighbours when $\langle \Delta x^2(l^*) \rangle = a_0^2$, where a_0 is the vortex spacing. The effective value of l that enters $U_p(l)$ is thus $l^* = l_c(a_0/\xi_\perp)^{1/\zeta} \propto B^{-3/2}$, where $B = \phi_0/da_0$ is the field associated with a planar array of vortices. On balancing the energy gain per unit area above H_{c1} of $N = W/a_0$ wandering flux lines, $g(H - H_{c1})/H_{c1}a_0 \equiv gha_0$, against the pinning energy density $U(l^*)/l^*a_0$ lost because of collisions between lines, we find the constitutive relation

$$B = (\phi_0/\xi_\perp d)(l_c h/\xi_\perp)^{\zeta/2(1-\zeta)} \equiv \frac{g^2 \xi_\perp \phi_0}{d\Delta} (H - H_{c1})/H_{c1}$$

for $\zeta = 2/3$, which is consistent with work on related problems^{16,17} with point disorder in 1 + 1 dimensions. For an analogous power-law constitutive relation in 2 + 1 dimensions, see ref. 12. To test our result we have extracted $B(H)$ from the data in Fig. 3b and plotted it as the 'staircase' in that figure. The data are a discrete approximation to the continuum prediction of a linear $B(H)$ near H_{c1} , thus confirming the anomalous wandering exponent $\zeta = 2/3$. This result is in striking contrast to the vertical slope predicted by Abrikosov's original theory of perfectly straight vortices in clean samples with exponential repulsion. From the slope of this linear dependence we obtain a simple measure of Δ and confirm that $\xi_\perp > l_c$.

Qualitative arguments based on this simple model suggest that relaxation times decrease markedly as the number of collisions grows with increasing B , leading to reversible behaviour for large B with residual fluctuations $\langle \delta B^2 \rangle \propto B^{-1/4}$. We expect that fluctuations in this regime provide a 'fingerprint' of the disorder in the sample. Universal sample-to-sample fluctuations have in fact been predicted for large planar assemblies of vortex lines strongly pinned by point disorder^{18,19}.

The study of the 1 + 1-dimensional random potential has been an important starting point in theoretical statistical physics and its investigation of disorder-dominated systems. This experiment is the first that directly studies this famous problem. We believe that the study of mesoscopic vortex physics might become an important new area of research for condensed-matter scientists. □

Received 2 December 1998; accepted 26 February 1999.

- Safar, H. *et al.* Experimental evidence for a 1st-order vortex-lattice-melting transition in untwinned single-crystal $\text{YBa}_2\text{Cu}_3\text{O}_7$. *Phys. Rev. Lett.* **69**, 824–827 (1992).
- Brézin, E., Nelson, D. R. & Thiaville, A. Fluctuation effects near H_{c2} in type-II superconductors. *Phys. Rev. B* **31**, 7124–7132 (1985).
- Kleiman, R. N., Agnolet, G. & Bishop, D. J. Two-level systems observed in the mechanical properties of single-crystal silicon at low temperatures. *Phys. Rev. Lett.* **59**, 2079–2082 (1987).
- Kleiman, R. N., Kaminsky, G. K., Reppey, J. D., Pindak, R. & Bishop, D. J. Single-crystal silicon high-Q torsional oscillators. *Rev. Sci. Instrum.* **56**, 2088–2091 (1985).
- Bishop, D. J. & Reppey, J. D. Study of the superfluid transition in two-dimensional ^4He films. *Phys. Rev. Lett.* **40**, 1727–1730 (1978).
- Bolle, C. A., de la Cruz, F., Gammel, P. L., Waszczak, J. V. & Bishop, D. J. Observation of tilt induced orientational order in the magnetic flux lattice in 2H-NbSe₂. *Phys. Rev. Lett.* **71**, 4039–4042 (1993).
- Duarte, A. *et al.* Dynamically induced disorder in the vortex lattice of 2H-NbSe₂. *Phys. Rev. B* **53**, 11336–11339 (1996).
- Yaron, U. *et al.* Neutron diffraction studies of flowing and pinned magnetic flux lattices in 2H-NbSe₂. *Phys. Rev. Lett.* **73**, 2748–2751 (1994).
- Akkermans, E., Montambaux, Pichard, J. I. & Zinn-Justin, J. (eds) *Mesoscopic Quantum Physics* (North-Holland, Amsterdam, 1994).
- de Trey, P., Gygax, S. & Jan, J.-P. Anisotropy of the Ginzburg–Landau parameter κ in NbSe₂. *J. Low Temp. Phys.* **11**, 421–434 (1973).
- Nelson, D. R. Vortex entanglement in high- T_c superconductors. *Phys. Rev. Lett.* **60**, 1973–1976 (1988).
- Natterman, T. & Lipowsky, R. Comment on vortex entanglement in high- T_c superconductors. *Phys. Rev. Lett.* **61**, 2508 (1988).
- Huse, D. A., Henley, C. L. & Fisher, D. S. Response on comment on roughening by impurities at finite temperature. *Phys. Rev. Lett.* **55**, 2924 (1985).
- Forster, D., Nelson, D. R. & Stephen, M. J. Large-distance and long-time properties of a randomly stirred fluid. *Phys. Rev. A* **16**, 732–749 (1977).
- Halpin-Healy, T. & Zhang, Y. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. *Phys. Rep.* **254**, 215–415 (1995).
- Hwa, T., Nelson, D. R. & Vinokur, V. M. Flux-line pinning by competing disorders. *Phys. Rev. B* **48**, 1167–1174 (1993).
- Hwa, T. & Lässig, M. Similarity detection and localization. *Phys. Rev. Lett.* **76**, 2591–2594 (1996).
- Hwa, T. & Fisher, D. S. Vortex glass phase and universal susceptibility variations in planar arrays of flux lines. *Phys. Rev. Lett.* **72**, 2466–2469 (1994).
- Fisher, M. P. A. Vortex-glass superconductivity: a possible new phase in bulk high- T_c oxides. *Phys. Rev. Lett.* **62**, 1415–1418 (1989).

Acknowledgements. We thank J. Blatter, T. Hwa, V. Kogan and V. Vinokur for helpful discussions.

Correspondence and requests for materials should be addressed to D.J.B. (e-mail: djb@lucent.com).

Influence of a knot on the strength of a polymer strand

A. Marco Saitta*, Paul D. Soper†, E. Wasserman† & Michael L. Klein*

* Center for Molecular Modeling, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6202, USA

† DuPont Central Research and Development, Expt. Station, Wilmington, Delaware 19880-0328, USA

Many experiments have been done to determine the relative strengths of different knots, and these show that the break in a knotted rope almost invariably occurs at the point just outside the 'entrance' to the knot¹. The influence of knots on the properties of polymers has become of great interest, in part because of their effect on mechanical properties². Knot theory^{3,4} applied to the topology of macromolecules^{5–8} indicates that the simple trefoil or 'overhand' knot is likely to be present in any long polymer strand^{9–12}. Fragments of DNA have been observed to contain such knots in experiments^{13,14} and computer simulations¹⁵. Here we use *ab initio* computational methods¹⁶ to investigate the effect of a trefoil knot on the breaking strength of a polymer strand. We find that the knot weakens the strand significantly, and that, like a knotted rope, it breaks under tension at the entrance to the knot.

Little is known about the structure and properties of knots at the atomic level. For example, the minimum number of carbon atoms that can be sustained as a trefoil in a polyethylene strand is not well established^{5,19}. Polyethylene is the simplest polymer and therefore an excellent generic system to study the fundamental properties of a knotted chain. (Rigorously, a knot is a closed loop; but here, as in ref. 1, we apply the term to a strand with unlinked ends.) As a representative polyethylene-like system we chose the linear molecule *n*-decane ($\text{C}_{10}\text{H}_{22}$). Empirical models that utilize force constants for bond strength, bend and torsion^{18–20}, although useful for studying bulk properties of chain molecules, are not applicable to the case of chain rupture. First-principles calculations¹⁶, on the other hand, have been shown to yield satisfactory results for structural and mechanical properties of hydrocarbon-based polymeric systems^{21,22}. The *ab initio* equilibrium structural parameters for such alkanes, after optimization, are in excellent agreement with experiment, with errors in bond lengths and angles smaller than 1%

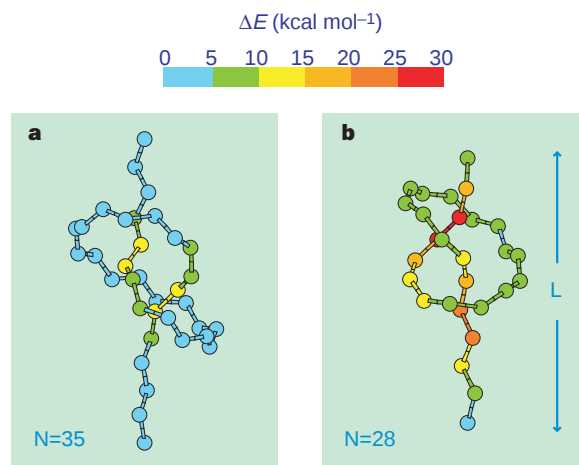


Figure 1 Strain energy distribution in a knotted polymer strand. Shown are distributions in chains of 35 (a) and 28 (b) carbon atoms taken from constrained classical MD simulations. When the knot is sufficiently tightened, the strain energy localizes most on the bonds immediately outside its entrance points.

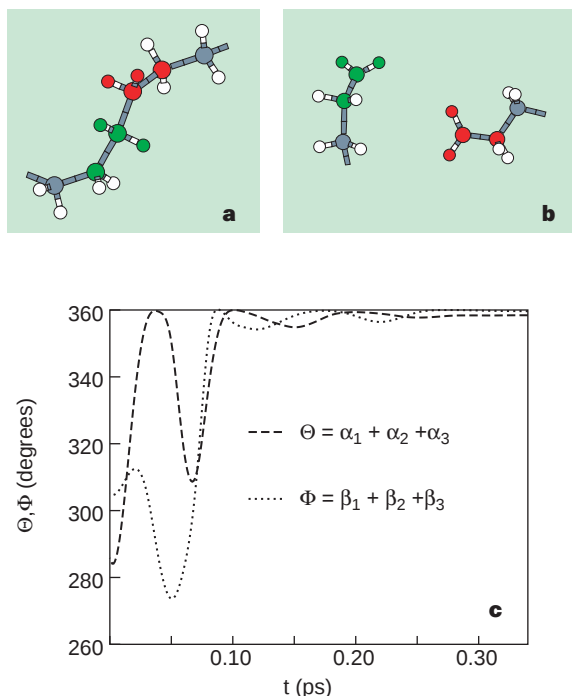


Figure 2 Analysis of chain rupture. The breakpoint in an $N = 28$ trefoil before (a) and after (b) chain dissociation, with C_{22} and C_{23} and their neighbouring atoms highlighted in green and red, respectively. Panel c shows the time evolution of the sum of the respective bond angles, which reveals an asymptotic sp^2 hybridization of the carbons involved in the break. The calculations are based on density-functional theory with Becke and Lee-Yang-Parr (BLYP)^{24,25} corrections to the exchange and correlation functionals, respectively. Electronic spin polarization has been included within the local spin-density approximation scheme. A valence electron pseudopotential scheme is employed with a plane-wave kinetic-energy cut-off at 60 Ry and a Γ -point sampling of the Brillouin zone. The simulation supercell is large enough to avoid interactions with periodic images.

(ref. 22). We studied the response of the n -decane molecule to uniaxial strain by systematically increasing the separation L between the terminal carbon atoms. For a fixed value of this separation, the system was heated to room temperature and then cooled via a simulated annealing technique. The decane molecule remained intact for an elongation up to 18.5% beyond its equilibrium length. Deviations of the C_1 – C_2 and C_9 – C_{10} bond lengths from the mean value increase significantly with the applied strain. Under maximum loading, the terminal bonds become stretched to ~ 1.8 Å compared with ~ 1.7 Å for the others and the terminal C–C–C bond angles increase to $\sim 135^\circ$.

For larger elongation, the molecule dissociates into two radicals with $C_{10}H_{22} \rightarrow C_9H_{19}^\bullet + CH_3^\bullet$. The bond that dissociates involves (randomly) one of the two atoms where the tension is applied. We obtain 83 kcal mol^{−1} for the dissociation energy for the methyl group, which compares favourably with the experimental dissociation enthalpy of ~ 87 kcal mol^{−1}. A similar calculation performed for n -undecane, $C_{11}H_{24}$, but with the tension still applied to the C_1 and C_{10} atoms, shows that the C_9 – C_{10} bond breaks to yield an ethyl radical. The calculated dissociation energy in this case is 81 kcal mol^{−1} compared with the experimental enthalpy of ~ 82 kcal mol^{−1}. As in the previous case, the hybridization of the two key carbon atoms changes from the initial tetrahedral sp^3 to planar sp^2 on formation of the product radicals.

To generate a starting configuration for the *ab initio* study we first performed classical molecular dynamics (MD) calculations on a loosely knotted polyethylene-like alkane chain, with $N = 144$ carbon atoms, in which a trefoil was introduced by adding an

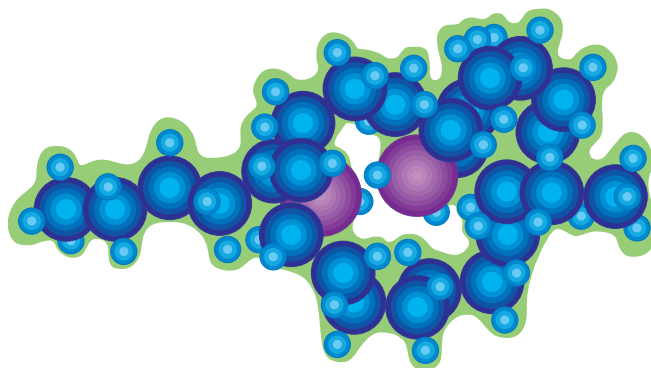


Figure 3 Schematic electron charge density immediately after the break. Carbon and hydrogen atoms are displayed with spheres corresponding to their respective covalent radii. The green area is a contour plot of the region containing 87% of the total electronic charge. A gap in the charge density, and thus bond-breaking, is observable between the two highlighted C atoms.

appropriate set of *gauche* defects¹⁷. The potential model employed was based on the united atom scheme^{18,19} but with *ab initio* intra-chain stretch and bend force constants taken from the present study. MD calculations were performed on the knotted chain for a sequence of increasing (constrained) end-to-end distances, L , which gradually resulted in a tighter knot. For any given L , 200 ps of room-temperature dynamics, followed by 50 ps of slow cooling, was utilized to relax the system in the trefoil configuration. (This run length seemed appropriate given that even the slowest longitudinal phonons can propagate along the entire length of the molecule about 1,000 times during this trajectory.)

The global (either constrained or unconstrained) minimum corresponds to an unknotted polymer, but the migration barrier for the knot to propagate along the chain, and thus the disentanglement time, increases dramatically as the trefoil tightens. Thus even at room temperature the system is ‘trapped’ in the knotted configuration for the timescale of the calculations. Under these circumstances, the portions of the chain far from the trefoil, although extremely important in disentanglement processes, are likely to have little or no effect on the behaviour of the knot under stress. Accordingly, as the knot was tightened, atoms far from it were removed from the chain, in order to eventually make the system tractable with first-principles methods.

For each classical MD calculation, we monitored the strain distribution along the chain. Even for $N = 50$, the chain with a trefoil stores a rather small amount of strain energy. But when the chain was shortened to $N = 35$ a pronounced inhomogeneity appeared in the strain energy distribution (Fig. 1a). The shape of the strain energy distribution is essentially the same, at a given L , independent of the details of the simulation, such as the initial configuration and temperature or the length of the MD simulation or a reasonable change in the interaction potentials. (The potentials that we employed in the MD simulations were all fitted to the properties of alkane molecules close to their equilibrium structures. Thus, in the present case where we are dealing with highly distorted systems, the model’s force constants are likely to yield only semi-quantitative results.) Further shortening of the chain to $N = 28$ produced a trefoil (Fig. 1b) with distortions in bond lengths and angles still below the critical values that yield bond breaking in the linear alkanes.

The constrained equilibrium positions obtained from the 250 ps of classical MD evolution of the C_{28} n -alkane provided the carbon skeleton of the starting configuration for the first-principles study of the $C_{28}H_{58}$ molecule. A series of Car-Parrinello MD (CPMD) simulations^{16,23–25} were performed with the separation between C_1 and C_{28} fixed at distances L between 11.50 and 14.00 Å. In each case, the system was allowed to evolve dynamically at room temperature,

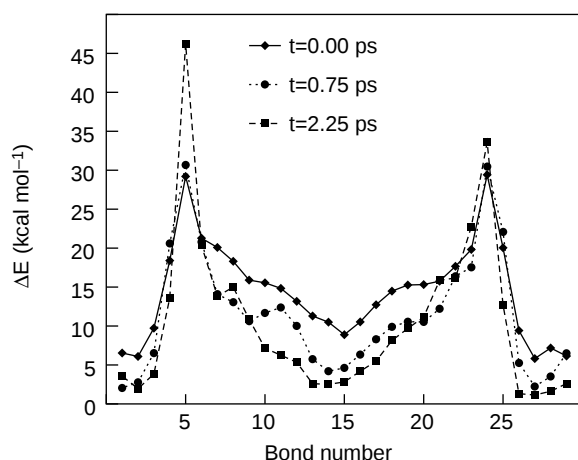


Figure 4 Relaxation of the strain energy distribution. Shown is the distribution along a $C_{30}H_{62}$ chain containing a trefoil during CPMD energy minimization at $T = 0\text{ K}$. The $t = 0$ configuration is an average of the equilibrium positions obtained from classical MD simulations performed with different initial states and temperatures. The segments outside the knot and its central portion relax, while stress tends to concentrate mostly on the two entrance bonds (see Fig. 1).

after which it was cooled. Typically, during the dynamical evolution individual bonds undergo large-amplitude thermal fluctuations, which decrease on cooling.

During the room temperature $L = 13.50\text{ \AA}$ CPMD run for $N = 28$, dissociation occurred at a bond location just outside the entrance to the knot, well separated from the terminal atom where the tension was applied. Figure 2 shows the region of the knot where the chain actually ruptured, both before (Fig. 2a) and after (Fig. 2b) the break. The change in hybridization of the C_{22} and C_{23} atoms from sp^3 to sp^2 is clearly evident also in Fig. 2c, where we display the time evolution of the system immediately before and after the break. Figure 3 shows a 'snapshot' of the electronic charge density after the break occurred. The gap in the charge density between the C_{22} and C_{23} atoms confirms the bond breaking at this position. Other simulations carried out with $L \geq 13.50\text{ \AA}$ yield breakpoints at one of the two bonds coming out from the knot. (Density-functional energy barriers are not quantitatively very accurate, but the general qualitative description of phenomena, as well as the dissociation energies involved in such processes, are known to be quite reliable.)

The first-principles CPMD calculations provide the total strain energy but not its profile along the chain. The latter can be obtained by making use of an *ab initio* force-constants model, which is able to reproduce the total strain energy to $\sim 20\%$. Figure 4 shows the evolution of the strain energy distribution during structural relaxation of a $N = 30$ trefoil just before breaking occurs. The strain energy is mainly localized in the two symmetric bonds that are outside the entrance to the knot. Our calculations suggest that 23 carbon atoms form the tightest knot that can be sustained in a polyethylene strand without it breaking.

The strain energy stored in the knot at breaking point is $12.7\text{ kcal mol}^{-1}$ per C–C bond, which is considerably smaller than the value ($16.2\text{ kcal mol}^{-1}$) for the linear unknotted case. Thus, we find that the presence of the knot has significantly weakened the strand in which it is tied.

Although the dynamical evolution of the present constrained *ab initio* MD simulations was sufficient to observe bond breaking, no attempt was made to allow for recombination of the resulting radicals or reactions of the radicals with other parts of the chain. The study of these effects, as well as the role of chain branching and the influence of neighbouring chains, is left for future research. These factors will probably provide a deeper understanding of how the interplay between inter- and intra-molecular effects contributes to

the mechanical properties of real polymer samples².

Note added in proof: Arai *et al.*²⁶ have recently reported the knotting of actin filaments and DNA molecules using optical tweezers. They find that the breaking stress for actin is significantly lower than that of the unknotted filaments, and that the breakage point is at the entrance to the knot, as our calculations predict. □

Received 22 September 1998; accepted 22 February 1999.

- Ashley, C. W. *The Ashley Book of Knots* (Doubleday, New York, 1993).
- Bayer, R. K. Structure transfer from a polymeric melt to the solid state. Part III: Influence of knots on structure and mechanical properties of semicrystalline polymers. *Colloid Polym. Sci.* **272**, 910–932 (1994).
- Atiyah, M. F. *The Geometry and Physics of Knots* (Cambridge Univ. Press, 1990).
- Katritch, V. *et al.* Geometry and physics of knots. *Nature* **383**, 142–145 (1996); Properties of ideal composite knots. *Nature* **388**, 148–151 (1997).
- Frisch, H. L. & Wasserman, E. Chemical topology. *J. Am. Chem. Soc.* **83**, 3789–3795 (1961).
- Mislow, K. *Introduction to Stereochemistry* (Benjamin, New York, 1965).
- Schill, G. *Catenanes, Rotaxanes, and Knots* (Academic, New York, 1971).
- Walba, D. M. Topological stereochemistry. *Tetrahedron* **41**, 3161–3212 (1985).
- Frank-Kamenetskii, M. D., Lukashin, A. V. & Vologodskii, A. V. Statistical mechanics and topology of polymer chains. *Nature* **258**, 398–402 (1975).
- Frisch, H. L. Macromolecular topology—Metastable isomers from pseudo interpenetrating polymer networks. *New J. Chem.* **17**, 697–701 (1993).
- Mansfield, M. L. Knots in hamiltonian cycles. *Macromolecules* **27**, 5924–5926 (1994).
- van Rensburg, E. J., Summers, D. A. W., Wasserman, E. & Whittington, S. G. Entanglement complexity of self-avoiding walks. *J. Phys. A* **25**, 6557–6566 (1992).
- Wasserman, S. A. & Cozzarelli, N. R. Biochemical topology: applications to DNA recombination and replication. *Science* **232**, 951–960 (1986).
- Shaw, S. Y. & Wang, J. C. Knotting of a DNA chain during ring closure. *Science* **260**, 533–536 (1993).
- Schlick, T. & Olson, W. K. Trefoil knotting revealed by molecular dynamics simulations of supercoiled DNA. *Science* **257**, 1110–1115 (1992).
- Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
- de Gennes, P.-G. Tight knots. *Macromolecules* **17**, 703–704 (1984).
- Siepmann, J. I., Karaborni, S. & Smit, B. Simulating the critical-behaviour of complex fluids. *Nature* **365**, 330–332 (1993).
- Mundy, C. J., Balasubramanian, S., Bagchi, K., Siepmann, J. I. & Klein, M. L. Equilibrium and non-equilibrium simulation studies of fluid alkanes in bulk and at interfaces. *Faraday Discuss.* **104**, 17–36 (1996).
- Karasawa, N., Dasgupta, S. & Goddard, W. A. Mechanical properties and force-field parameters for polyethylene crystal. *J. Phys. Chem. US* **95**, 2260–2272 (1991).
- Ancilotto, F., Chiarotti, G. L., Scandolo, S. & Tosatti, E. Dissociation of methane into hydrocarbons at extreme (planetary) pressure and temperature. *Science* **275**, 1288–1290 (1997).
- Montanari, B. & Jones, R. O. Density functional study of crystalline polyethylene. *Chem. Phys. Lett.* **272**, 347–352 (1997).
- Martyna, G. J. *et al.* PINY Code. *Comp. Phys. Comm.* (in the press).
- Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
- Lee, C., Chang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
- Arai, Y. *et al.* Tying a molecular knot with optical tweezers. *Nature* (in the press).

Acknowledgements. This work was supported in part by the National Science Foundation.

Correspondence and requests for material should be addressed to M.L.K. (e-mail: klein@lrsu.upenn.edu).

Controlled growth and electrical properties of heterojunctions of carbon nanotubes and silicon nanowires

Jiangtao Hu*†, Min Ouyang*†, Peidong Yang‡ & Charles M. Lieber*§

* Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA

‡ Department of Chemistry, University of California, Santa Barbara, California 93106, USA

§ Division of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts 02138, USA

† These authors contributed equally to this work.

Nanometre-scale electronic structures are of both fundamental and technological interest: they provide a link between molecular and solid state physics, and have the potential to reach far higher

device densities than is possible with conventional semiconductor technology^{1,2}. Examples of such structures include quantum dots, which can function as single-electron transistors^{3,4} (although their sensitivity to individual stray charges might make them unsuitable for large-scale devices) and semiconducting carbon nanotubes several hundred nanometres in length, which have been used to create a field-effect transistor⁵. Much smaller devices could be made by joining two nanotubes or nanowires to create, for example, metal–semiconductor junctions, in which the junction area would be about 1 nm² for single-walled carbon nanotubes. Electrical measurements of nanotube ‘mats’ have shown the behaviour expected for a metal–semiconductor junction⁶. However, proposed nanotube junction structures⁷ have not been explicitly observed, nor have methods been developed to prepare them. Here we report controlled, catalytic growth of metal–semiconductor junctions between carbon nanotubes and silicon nanowires, and show that these junctions exhibit reproducible rectifying behaviour.

Two approaches were used to prepare nanotube–nanowire (NT/NW) junctions (Fig. 1). These methods built on recent efforts in our group⁸ and elsewhere^{9,10} directed towards the controlled growth of NWs and NTs with catalysts. It is possible to extend the basic ideas from this previous work to create NT/NW junctions by localizing a suitable metal catalyst at the end of a starting NW or NT. We have met this critical requirement by using a common catalyst for the growth of the NW and NT, and by depositing a NW catalyst at the end of an existing NT.

A common Fe-based catalyst was used to grow NTs from the ends of silicon NWs (SiNWs). In previous work, we have shown that Fe and Fe/Au are efficient catalysts for the growth of SiNWs⁸; moreover, it is well established that Fe is a good catalyst for the growth of multiwalled NTs (MWNTs)⁹ and single-walled NTs (SWNTs)¹⁰. The use of a common catalyst has the advantage that the catalyst is naturally localized at the ends of the SiNWs after their growth, and then can be used to direct the growth of NTs using a hydrocarbon reactant such as ethylene (Fig. 1a). The SiNWs grown with Fe/Au catalysts are single-crystal materials with an average diameter of ~10 nm (Fig. 2a). Transmission electron microscopy (TEM) and X-ray fluorescence (EDX) analyses further demonstrate that the SiNWs terminate in well-defined Fe/Au nanoclusters.

NTs were grown from the ends of the SiNWs by using ethylene at 600 °C. TEM studies of the products produced in this way show a large number of NT/SiNW junctions (for example, Figs 2b, c), where NTs are clearly hollow and SiNWs are solid. Electron diffraction (Fig. 2b, inset) and high-resolution imaging reveal that the SiNWs remain crystalline after NT growth. The observed NT wall thicknesses and lattice-resolved imaging demonstrate that the NTs are well-ordered MWNTs as expected for these growth

conditions⁹. We find that two general junction structures are produced from the SiNW ends. In the more common case, the NT forms a sharp junction with the SiNW (Fig. 2b). A metal nanocluster, which can be excluded during growth, is often observed nearby the junction. Because MWNTs are typically metallic¹¹, these sharp NT/SiNW nanojunctions can exhibit behaviour typical of metal–semiconductor (M/S) junctions formed on planar silicon¹². In the second less frequent type of junction, a catalyst cluster remains in the region between the NT and SiNW (Fig. 2c, large arrow). This latter image also exhibits contrast changes in the SiNW portion of the junction. EDX and electron diffraction measurements demonstrate that these regions do not contain significant metal catalyst

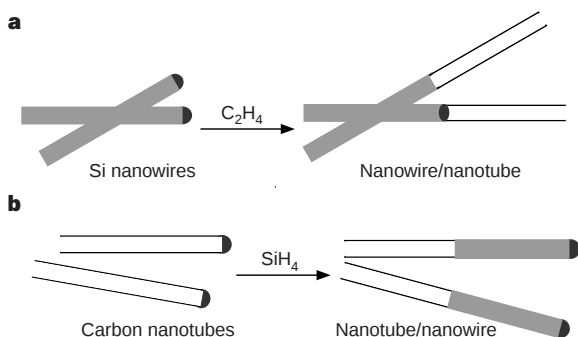


Figure 1 Synthetic approaches to NT/SiNW junctions. **a**, SiNWs (grey) grown by a catalytic process⁸ terminate in nanocluster catalysts (black). This catalyst is used to direct the growth of NTs (black lines) from ethylene. The catalyst can be excluded or remain at the NT/SiNW junction. **b**, Catalyst nanoclusters (black) are deposited on NT ends and then used to direct the growth of SiNWs. In this case, the catalyst is on the SiNW free end and the NT/SiNW junction is clean.

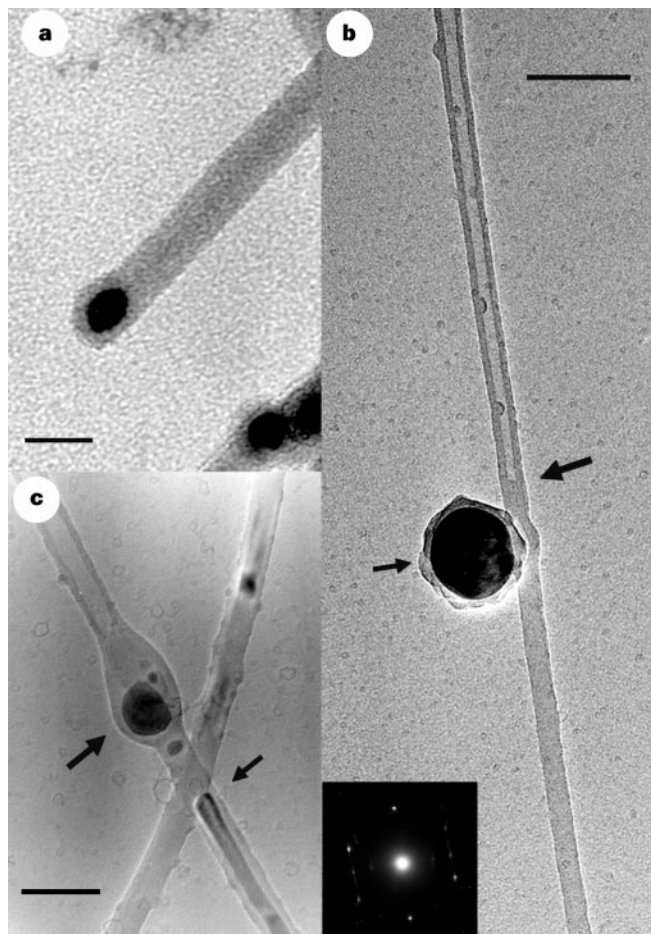


Figure 2 NT/SiNW junctions grown from SiNWs with a common Fe/Au catalyst. **a**, TEM image (EM420; Philips) of SiNW produced with a Fe_{0.9}Au_{0.1} catalyst. The 15-nm-diameter NW terminates in a Fe–Au nanocluster that appears as a dark, solid sphere. **b**, **c**, TEM images of NT/SiNW junctions. The hollow NTs are oriented at the top of both images and the solid SiNWs at the bottom. **b**, The large black arrow highlights the junction position. The small arrow indicates a metal catalyst cluster that was probably excluded from the junction during growth. The inset corresponds to an electron diffraction pattern recorded with the beam perpendicular to the NT/SiNW axis. The diffraction spots show that the SiNW maintains its crystalline structure after NT growth. **c**, The larger arrow highlights the junction region that contains a metal nanocluster. The contrast variation highlighted by the smaller arrow corresponds to SiNW (see the text for details). The scale bars in **a**, **b** and **c** are 20, 50 and 50 nm, respectively. Samples were prepared as follows. A pulsed Nd–YAG laser (532 nm, 7-ns pulse width; GCR-16S; Spectra-Physics) was used to ablate a Fe_{0.9}Au_{0.1} solid target placed within a growth system⁸. The ablation was performed for 10 min in 20-torr 10%/90% silane/He (99.998%; Matheson Gas Products), 20 standard cm³ min^{−1} at 450 °C. NTs were grown from the SiNWs for 30 min in 300-torr ethylene (99.5%; Matheson Gas Products), 20 standard cm³ min^{−1} at 600 °C.

and are crystalline Si. We believe that the contrast changes are due to Si recrystallization and texturing mediated by Au(Fe) catalyst during NT growth. This type of junction is more complex than the more common sharp NT/SiNW structure but should also exhibit M/S behaviour, as both the NT and catalyst particles are metallic.

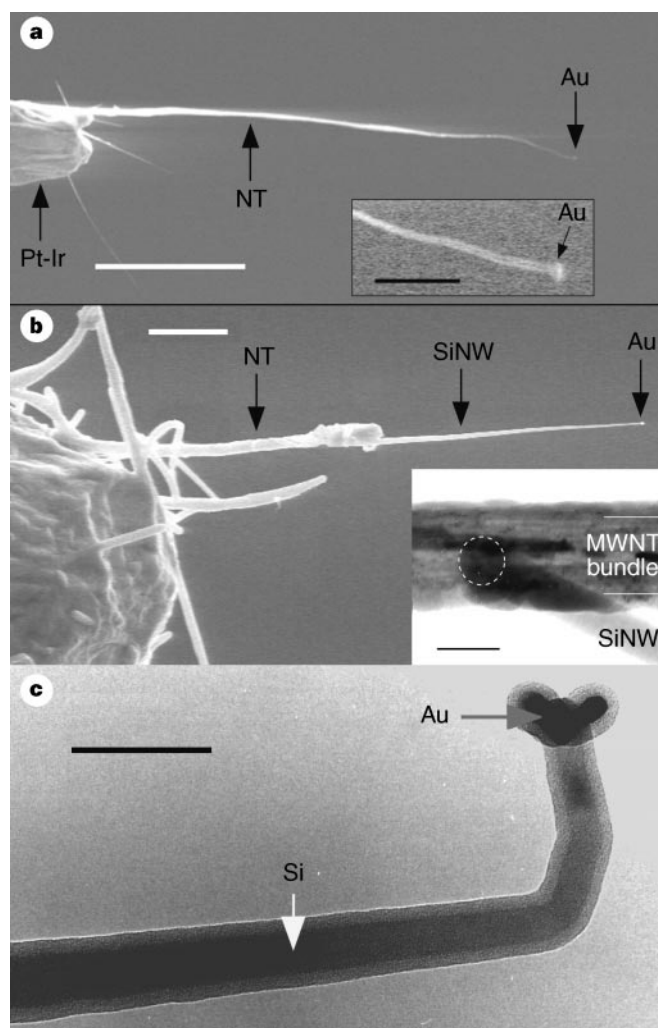


Figure 3 NT/SiNW junctions grown from NT tips with a Au catalyst. **a**, FESEM images (Leo 982; Leo Electron Microscopy) of a NT with an electrodeposited Au nanocluster at the free end. The NT is attached to a Pt-Ir STM tip with micromanipulators¹³. Inset, higher-magnification view of the Au nanocluster at the NT end. The apparent doubling of the NT is due to vibration of the long tube. The white and black scale bars correspond to 5 μm and 500 nm, respectively. **b**, FESEM image of the NT/SiNW junction grown from a nanotube tip. The large structure at the left is the Pt-Ir tip; several spurious nanotubes are present on the tip surface from the initial transfer process. Scale bar, 2 μm . Inset, TEM image of a NT/SiNW junction. The junction, which is indicated by the dashed white circle, is located at the back of the MWNT bundle and seems clean (for example, there is no evidence for Au in the image). The extent of the MWNT bundle is indicated with white lines; amorphous Si/SiO_x covers the surface as described in the text. Scale bar, 50 nm. **c**, TEM image of the tip of the nanowire in **b**. The nanowire consists of a crystalline core and an amorphous Si/SiO_x coating. Diffraction and EDX studies demonstrate that there is undetectable (<0.5%) gold in the nanowire and NT/SiNW junction, and that there is a well-defined Au nanocluster at the SiNW free end. Scale bar, 50 nm. Samples were prepared as follows. MWNTs¹⁸ were mounted onto freshly cut Pt-Ir STM tips with silver paste (H20E; Epoxy Technology), and gold nanoclusters were electroplated on the free NT ends (Techni-Gold 25 E; Technic) at 500 mV. SiNWs were then grown from the NT ends by chemical vapour deposition (CVD) (10%/90% silane/He, 20 standard cm³ min⁻¹ for 15 min at 510 °C).

We have also studied growth from catalysts attached to NTs. This method provides greater flexibility in the choice of catalyst and enables direct electrical measurements, as NTs can be easily attached to scanned probe microscopy tips¹³. NTs were attached to sharpened Pt-Ir STM tips, and then a gold cluster was electrodeposited onto the NT free ends. Typical field-emission scanning electron microscope (FESEM) images of an attached MWNT bundle show that it is possible to localize Au nanoclusters at the very end of NT tips (Fig. 3a). The apparent splitting in the high-resolution image (Fig. 3a, inset) is due to the vibration of the long NT.

To create NT/SiNW junctions, SiNWs were grown from the NT tip ends by using silane¹⁴. A typical FESEM image of a NT/SiNW junction (Fig. 3b) shows roughly equal 5 μm straight segments of NT and SiNW with the junction near the centre of the image. TEM and FESEM images of the SiNW portion of the structure further demonstrate (1) that the NWs consist of a crystalline Si core sheathed in a thin, amorphous Si (α -Si) and native SiO_x layer (which forms after the sample is exposed to air), (2) that the Au catalyst used to direct growth is at the SiNW end (Fig. 3c), and (3) that the nanotube and SiNW can be either on-axis (Fig. 3b) or off-axis (Figs 3b and 4a, insets). TEM and EDX also demonstrate that the NT/SiNW structures prepared in this way do not have detectable Au, α -Si or SiO_x at the junction interface (Fig. 3b, inset). We believe that the preference for growth with Au catalyst at the SiNW free end reflects the more favourable energetics of NT–solid Si versus NT–liquid Au/Si interfaces.

There are two other features of the NT/SiNW junctions produced by this latter method that are evident in Fig. 3. First, the NT diameters are larger after SiNW growth. This reflects the uncata-

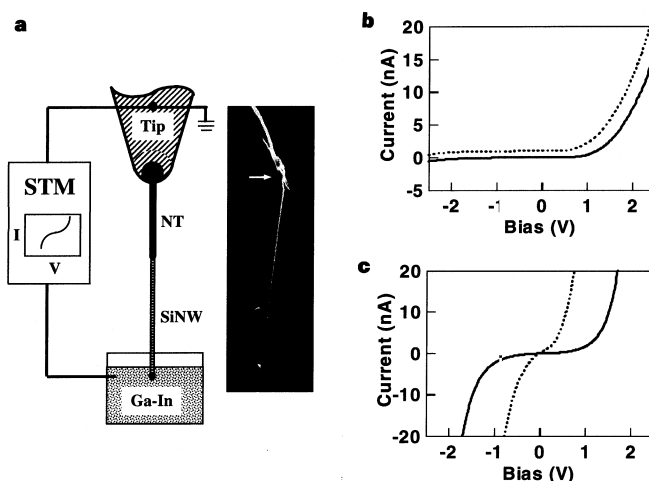


Figure 4 Electrical properties of NT/SiNW junctions. **a**, Diagram of the set-up used to measure the electrical properties of nanojunctions. The NT/SiNW is electrically connected to a Pt-Ir tip with Ag paste at the NT end (filled black circle) and to Ga-In liquid metal at the SiNW end. *I*-*V* curves are measured with the scanning tunnelling microscope (STM) electronics. Inset, another NT/SiNW junction grown as described in Fig. 3. The arrow indicates the position of the junction with the NT and SiNW above and below this point, respectively. **b**, *I*-*V* curves of two independent NT/SiNW junctions (solid and dotted curves). The reverse bias region of the dotted curve overlaps the solid curve and was shifted for clarity. **c**, *I*-*V* curves of a NT (dotted), coated with a thin layer of amorphous silicon, and a SiNW (solid) connected directly to the Pt-Ir tip. The current through SiNWs was 3–5 times that through NT/SiNW junctions over the voltage range used to determine ϕ_b . This shows that forward bias response of the NT/SiNW structures is dominated by the internal junction. Experimental procedures were as follows. Pt-Ir tips with NT/NW junctions were mounted in a commercial STM (Nanoscope III; Digital Instruments) equipped with a variable-gain current amplifier (QME 311; Balzers). The NT/SiNW junctions were advanced or retracted in 25 nm steps into or from the Ga-In liquid (99.99%; Alfa Aesar), with the STM feedback loop disabled; *I*-*V* measurements were made at a variety of depths.

lysed deposition of α -Si (from silane) during NW growth. Second, many of the SiNWs grown on NTs with the Au catalyst seem to taper towards the end. A reduction in diameter with increasing length has been reported previously for larger Si whiskers; this reduction was attributed to the incorporation of gold in the Si phase during growth¹⁴. Although we do not have direct evidence for Au-doping of the SiNWs, it should be noted that Au is expected to create deep acceptor and donor levels in Si (ref. 15) that are electrically inactive.

We have measured directly the electrical properties of individual NT/SiNW junctions grown from NT tips with an approach that exploits a liquid metal¹⁶ to make the second electrical contact to the free ends of NT/SiNW junctions (Fig. 4a). The current–voltage (I – V) measurements on NT/SiNW junctions with only the SiNW in contact with the Ga–In liquid exhibit very reproducible rectifying behaviour typified by the results for two independent junctions shown in Fig. 4b. Specifically, there is little current flow for $V < 0$, whereas the current increases sharply for voltages $\approx +0.8$. Similar results were obtained in nine independent nanojunctions. Significantly, this rectifying behaviour is characteristic of a M/S Schottky diode device.

Several experiments lead us to conclude that this diode-like behaviour is intrinsic to the nanoscale NT/SiNW junctions. First, I – V curves remain unchanged as the tip is advanced into the Ga–In liquid as long as this distance does not exceed the measured length of the SiNW. These results suggest that the overall current flow is limited by the actual NT/SiNW junction. Second, when the tips are advanced past the SiNW so that the NT is in contact with Ga–In liquid, symmetrical, non-rectifying I – V curves are observed (Fig. 4c). These results show that the NT/Pt–Ir junction does not produce the observed diode characteristics. The non-ohmic I – V curves recorded when the NT is in contact with Ga–In are due to the barrier created by the thin α -Si/SiO_x coating formed on the NT during NW growth and subsequent exposure to air (see above). Third, I – V measurements made on SiNWs grown directly from the Pt–Ir tips (the same conditions as for NT/SiNW growth) are symmetrical, exhibit no evidence of rectifying behaviour but have characteristics of a tunnel junction (Fig. 4c). These results demonstrate that the observed diode behaviour is not due to the SiNW/Ga–In contact or a Si/ α -Si interface. We believe that the SiNW/Ga–In contact is a tunnel junction due to the native surface oxide layer¹², whereas the SiNW/Pt contact is ohmic, as observed previously for p-Si/Pt junctions¹⁷. Last, the nine independent NT/SiNW junctions that we have synthesized all exhibit the well-defined rectifying behaviour shown in Fig. 4b, thus testifying to the reproducibility of these nanojunctions.

The NT/SiNW M/S junctions exhibit Schottky diode behaviour similar to much larger planar Si-based structures, despite their small size. To explore the similarity further we have analysed the forward bias I – V data to extract the zero-bias junction barrier height, ϕ_b (ref. 12). It is interesting that this provides values of ϕ_b , 0.41 ± 0.06 eV, very similar to those obtained for macroscopic metal/p-Si junctions¹², in which we have verified that the SiNWs are p-doped (J.H., M.O. and C.M.L., unpublished observations).

Our controlled synthesis of NT/SiNW M/S junctions exhibiting Schottky diode behaviour opens up exciting opportunities on several fronts. They show that conventional semiconductor device structures of very small junction area can be embedded within NT/NW structures and that these quantum wires can serve as a natural means for electrically addressing the nanojunctions. Further studies of these structures, for example as a function of NW/NT diameter and temperature, will provide insight into the nanojunction physics and enable minimum size limits to be defined. More generally, we believe that the catalytic approach used to prepare NT/SiNW junctions can be readily expanded to create other M/S and semiconductor–semiconductor structures. Formation of junctions between p-type and n-type semiconductors could produce remarkably small light-emitting diodes that could also be readily addressed electrically. It is therefore reasonable to speculate that nanojunctions formed by the method described here could provide some of

the critical building blocks needed to enable the emerging technology of nanoelectronics. □

Received 23 November 1998; accepted 1 March 1999.

1. Carter, F. L. *Molecular Electronic Devices* (Marcel Dekker, New York, 1982).
2. Timp, G. *Nanoscience and Technology* (AIP Press, New York, in the press).
3. Devoret, M. H., Esteve, D. & Urbina, C. Single-electron transfer in metallic nanostructures. *Nature* **360**, 547–553 (1992).
4. Klein, D. L., Roth, R., Lim, A. K. L., Alivisatos, A. P. & McEuen, P. L. A single-electron transistor made from a cadmium selenide nanocrystal. *Nature* **389**, 699–701 (1997).
5. Tans, S. J., Verschuuren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
6. Collins, P. G., Zettl, A., Bando, H., Thess, A. & Smalley, R. E. Nanotube nanodevice. *Science* **278**, 100–103 (1997).
7. Chico, L., Crespi, V. H., Benedict, L. X., Louie, S. G. & Cohen, M. L. Pure carbon nanoscale devices: nanotube heterojunctions. *Phys. Rev. Lett.* **76**, 971–974 (1996).
8. Morales, A. & Lieber, C. M. A laser ablation method for the synthesis of crystalline semiconductor nanowires. *Science* **279**, 208–211 (1998).
9. Fonseca, A. *et al.* Synthesis of single- and multi-wall carbon nanotubes over supported catalysts. *Appl. Phys. A* **67**, 11–22 (1998).
10. Kong, J., Cassell, A. M. & Dai, H. Chemical vapor deposition of methane for single-walled carbon nanotubes. *Chem. Phys. Lett.* **292**, 567–574 (1998).
11. Dai, H., Wong, E. W. & Lieber, C. M. Probing electrical transport in nanomaterials: conductivity of individual carbon nanotubes. *Science* **272**, 523–526 (1996).
12. Rhoderick, E. H. *Metal–Semiconductor Contacts* (Clarendon, Oxford, 1978).
13. Dai, H., Hafner, J. H., Rinzler, A. G., Colbert, D. T. & Smalley, R. E. Nanotubes as nanopores in scanning probe microscopy. *Nature* **384**, 147–150 (1996).
14. Givargizov, E. I. in *Current Topics in Materials Science* (Kaldis, E. ed.) Vol. V1, 79–145 (North-Holland, New York, 1978).
15. Rao, K. S. R. K., Kumar, V., Premachandran, S. K. & Raghunath, K. P. Relationship of the gold related donor and acceptor levels in silicon. *J. Appl. Phys.* **69**, 2714–2716 (1991).
16. Frank, S., Poncharal, P., Wang, Z. L. & de Heer, W. A. Carbon nanotube quantum resistors. *Science* **280**, 1744–1746 (1998).
17. McCafferty, P. G., Sellai, A., Dawson, P. & Elabd, H. Barrier characteristics of PtSi/p-Si Schottky diodes as determined from I – V – T measurements. *Solid-State Electron.* **39**, 583–592 (1996).
18. Ebbesen, T. W. & Ajayan, P. M. Large-scale synthesis of carbon nanotubes. *Nature* **358**, 220–222 (1992).

Acknowledgements. We thank P. Kim and J. L. Huang for helpful discussions, and H. Wu and T. Deng for the Au electroplating solution. C.M.L. acknowledges support of this work by the ONR and NSF.

Correspondence and requests for materials should be addressed to C.M.L. (e-mail: cml@cmliris.harvard.edu).

Stress-induced recrystallization of a protein crystal by electron irradiation

F. Zemlin*, R. Schuster*, E. Beckmann*, J. L. Carrascosa†, J. M. Valpuesta† & G. Ertl*

* Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany

† Centro Nacional de Biotecnología, CSIC, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

Ordering of a system of particles into its thermodynamically stable state usually proceeds by thermally activated mass transport of its constituents. Particularly at low temperature, the activation barrier often hinders equilibration—this is what prevents a glass from crystallizing¹ and a pile of sand from flattening under gravity. But if the driving force for mass transport (that is, the excess energy of the system) is increased, the activation barrier can be overcome and structural changes are initiated². Here we report the reordering of radiation-damaged protein crystals under conditions where transport is initiated by stress rather than by thermal activation. After accumulating a certain density of radiation-induced defects during observation by transmission electron microscopy, the distorted crystal recrystallizes. The reordering is induced by stress caused by the defects at temperatures that are low enough to suppress diffusive mass transport. We propose that this defect-induced reordering might be a general phenomenon.

The protein structures we investigated were well-ordered crystals of the connector protein of the bacteriophage $\Phi 29$ (ref. 3). Crystals with a thickness of several monolayers and lateral dimensions of several micrometres were prepared as described in ref. 4. These

crystals were embedded in vitreous (amorphous) ice and held at a specimen temperature of 4.2 K in a cryo-electron microscope⁵.

Figure 1a shows a medium-resolution transmission electron microscope (TEM) image of a well-ordered multilayer crystal. Single molecules are arranged in a square structure with a spacing of 16.5 nm between the close-packed rows⁶. The weak additional feature in the centre of the unit cell is attributed to the three-dimensional structure of the protein crystal. The intensity autocorrelation of this image, reproduced in the inset, also exhibits a well-ordered square pattern. Cross-sections along the close-packed direction (Fig. 2a, trace A) show oscillations with almost constant amplitude, extending to large distances reflecting good long-range order of the crystal.

Owing to radiation damage by the electron radiation, the image fades after only a few seconds of irradiation (Fig. 1b). After about 14 s,

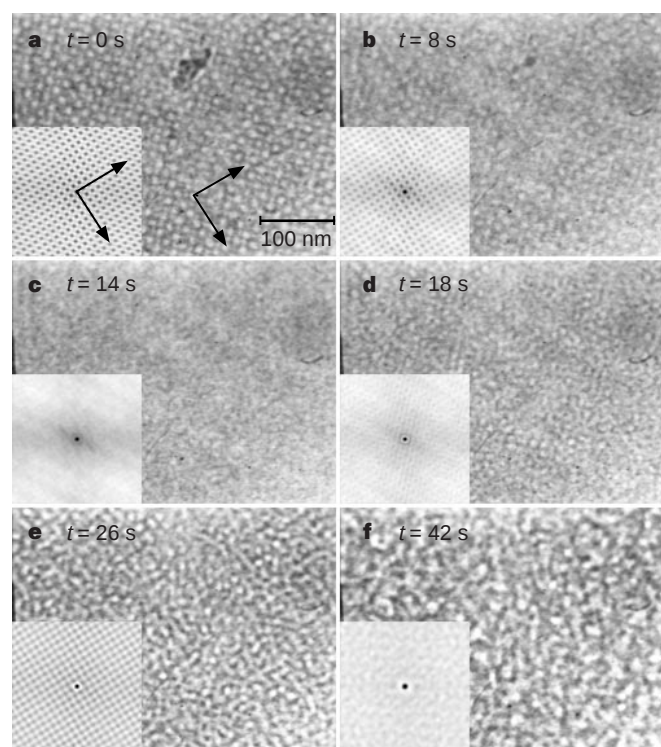


Figure 1 Sequence of TEM images of a protein crystal. The insets show the spatial autocorrelation of the TEM images. Irradiation times and the unit cell are indicated. The square lattice (a) fades (b, c) until the square structure re-evolves (d, e). After 42 s of irradiation, the periodic structure is irreversibly destroyed (f).

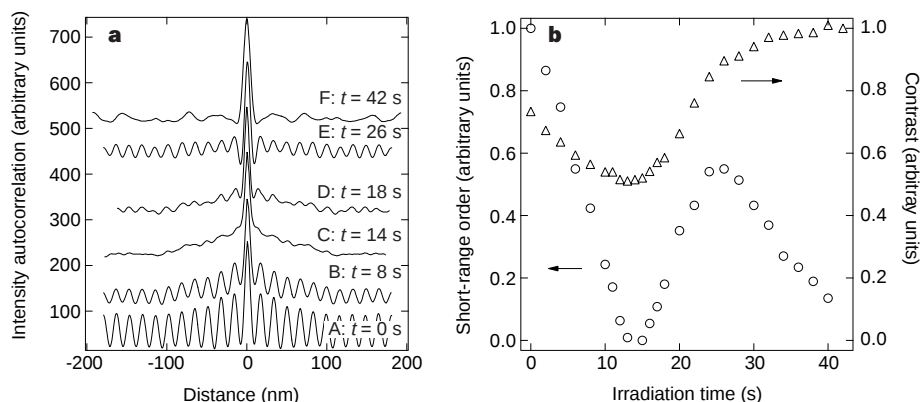


Figure 2 Time dependence of autocorrelation, short-range order and contrast of Fig. 1. a, Cross-sections through the spatial autocorrelation of the images in Fig. 1.

which corresponds to a dose of 10^4 electrons nm^{-2} , regular features can no longer be observed (Figs 1c and 2a, trace C). But strikingly, a few seconds later the contrast evolves again, as shown in the autocorrelation trace D in Fig. 2a. After 26 s of irradiation a square structure re-emerged (Fig. 1e), the autocorrelation of which exhibits long-range oscillations, perfectly in phase with those before irradiation (Fig. 2a, compare traces A and E). The partly destroyed crystal reorders with progressing time, maintaining its structure and orientation. Only upon further irradiation is this regular lattice destroyed irreversibly, leading to the formation of worm-like aggregates (Fig. 1f).

The temporal evolution of the degree of order upon irradiation is rationalized in Fig. 2b, where both the short-range order and the contrast of the TEM images are plotted against the time of irradiation. The short-range order corresponds to the height of the first maximum next to $\Delta x = 0$ in the intensity autocorrelation of the images. The contrast was obtained from the width of the greyscale distribution in the images. Upon irradiation, the contrast and the short-range order deteriorate within ~ 14 s. However, with progressing time, an almost stepwise increase of both image contrast and short-range order signals fast reordering of the protein crystal. Whereas the contrast stays approximately constant with further irradiation, the short-range order decreases continuously, signalling the proceeding radiation damage. These experiments have been repeated using different crystallites and samples, revealing analogous behaviour.

Figure 3 shows the typical evolution of the TEM diffraction patterns with progressing irradiation. In the beginning, sharp diffraction spots up to the second order indicate a well-ordered crystal. Equivalent to the results presented in Fig. 1, the diffraction pattern fades after 13 s of irradiation (Fig. 3b). After 21 s, sharp diffraction spots of the re-emerging crystal are clearly distinguishable (Fig. 3c), the distances and orientations of which coincide exactly with those at the beginning of the irradiation. Upon further irradiation the diffraction pattern fades again and, after about 40 s, only diffuse scattering is observable, forming a halo around the centre reflection.

It should be emphasized that the specimen holder was kept at 4.2 K in the cryo microscope. Embedding the protein crystals in vitreous ice and using carbon-covered copper TEM grids ensures that, even in the irradiated area of the crystal, the deposited energy is transferred efficiently as heat to the sample holder. Therefore, the temperature does not increase substantially above 20 K (refs 7, 8). The stability of protein crystals up to room temperature indicates strong intermolecular bonds. Thus, at low temperatures, thermally activated diffusion of the protein molecules is certainly suppressed and the motion and ordering of the molecules occur without surmounting a thermal activation barrier. Rather, the distortions of the lattice must have rendered it unstable against reordering, with stress as the driving force for mass transport.

The traces labelled A to F correspond to those in Fig. 1. b, Temporal evolution of the image contrast and the short-range order.

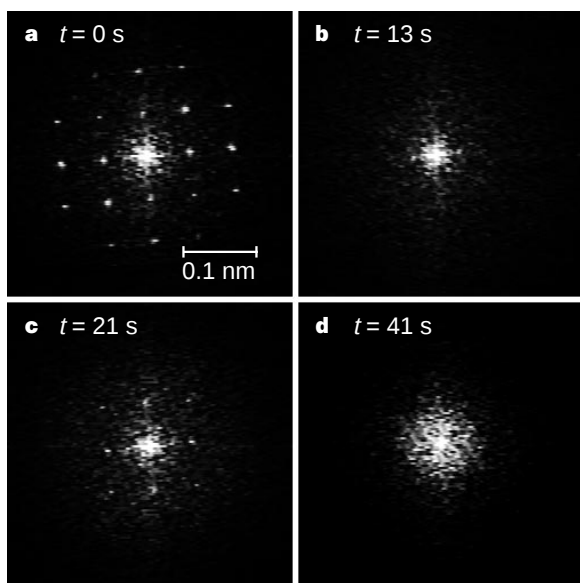


Figure 3 Electron diffraction images of a protein crystal under progressing irradiation. The irradiation times are indicated. Note that the faint hexagon around the centre spot, best visible in **b**, constitutes an artefact caused by the fibre optics of the image recording system.

The defects in the protein lattice are induced by electron radiation. Ionization events caused by the primary and secondary electrons lead to breakage of intramolecular bonds of the protein molecules. This results eventually in the formation of radicals, leaving the backbone of the protein molecule more-or-less intact⁹. In addition, upon neutralization, ionization energy of an order of several electron volts has to be dissipated in the crystal. This energy is certainly sufficient to break bonds between neighbouring protein molecules, thereby locally disordering the lattice. It is assumed that the lattice is distorted locally by shifting and rotating the molecules while the body of the proteins remains almost undamaged. At least during the first fading of the TEM image, no substantial chemical radiation damage is expected to take place. This is the reason why a crystal with the same lattice constant is restored after ~20 s of irradiation. Only at higher radiation doses do damaged molecules start to react and 'polymerize'. These products show up as worm-like aggregates at the end of the experiment, destroying the short-range order (Fig. 2b). Gas-bubble evolution by recombination of radiation products¹⁰ was observed only at ~10 times the dose for the reordering process.

Transport processes in the absence of thermal activation happen in a variety of physical systems. Among them are the avalanches in a pile of sand, triggered by a single additional grain of sand, as well as the magnetization of an impurity-containing material^{12–14}. Powerful theoretical concepts were developed to study systems in such a 'dynamical critical state'^{12–14}. However, in these systems the transport proceeds in an external field. But ordering in the absence of an external field is possible without thermal activation in systems other than the protein crystal studied here. A similar effect is demonstrated in Fig. 4, where we show the reordering of a distorted, defective soap-bubble crystal, produced by following the procedure of Bragg and Nye^{15,16}. In this model crystal, owing to the macroscopic size of its constituents, thermally activated processes are of no relevance.

Three domains of such an undistorted crystal are shown in Fig. 4a. By using an electrically heated needle, we randomly destroyed single bubbles of the middle crystallite (Fig. 4b). Although the creation of these point defects leads to considerable strain, the distorted structure remains stable over the timescale of the experiment, because a substantial barrier prevents the rearrangement of the bubbles: upon movement neighbouring bubbles would have

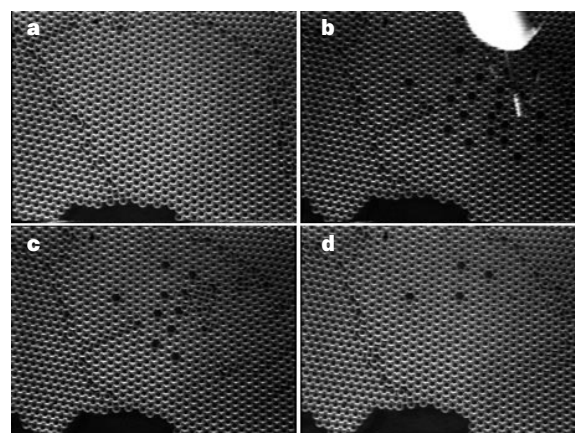


Figure 4 Recrystallization of a soap bubble 'crystal'. **a**, Three soap-bubble crystallites, consisting of bubbles of diameter 0.9 mm, floating on the soap solution. **b**, Several randomly chosen bubbles were destroyed with the needle imaged in the upper part of the image. **c**, After introducing 25 defects, the distorted lattice starts to reorder until the stable configuration is achieved.

to be displaced. Only if a critical density of defects is exceeded does reordering start without external influence. In Fig. 4c lower-coordinated bubbles start to move, finally eliminating 21 out of 25 defects (Fig. 4d). The induced strain in Fig. 4b was sufficient to overcome the barrier for bubble movement, and the remaining bubble network relaxes by reordering.

Note that irradiation of metal alloys at low temperatures, where the thermal diffusion of interstitial atoms is suppressed, may lead to amorphization, including substantial mass transport, driven by the disorder and strain¹⁷. Although the short-range order is re-established through this process, the long-range order vanishes. However, these effects were studied in large three-dimensional crystals, whereas both the protein and soap-bubble crystals are finite, allowing for the effective elimination of defects at the sample boundaries. This is presumably the reason why the stress-induced transport processes lead to crystalline order in the latter cases.

The above discussion implies that the observed stress-induced ordering is a rather general phenomenon that might be of particular relevance, for example, for the epitaxial growth of crystalline materials. Our experiments imply that in a disordered deposited layer, spontaneous reordering might occur at the point when the stress increases above a certain threshold during the growth. The stress necessary to induce the recrystallization certainly depends on the particular bond strength of the material. This may be the reason why such effects have not been reported before. □

Received 10 August 1998; accepted 9 March 1999.

1. Debenedetti, P. G. *Metastable Liquids: Concepts and Principles* (Princeton Univ. Press, Princeton, 1996).
2. Frette, V. et al. Avalanche dynamics in a pile of rice. *Nature* **379**, 49–52 (1996).
3. Valpuesta, J. M. & Carrascosa, J. L. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Q. Rev. Biophys.* **27**, 107–155 (1994).
4. Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. Analysis of electron microscope images and electron diffraction patterns of thin crystals of $\Phi 29$ connectors in ice. *J. Mol. Biol.* **240**, 281–287 (1994).
5. Zemlin, F., Beckmann, E. & Mast, K. D. v. d. A 200 kV electron microscope with Schottky field emitter and a helium-cooled superconducting objective lens. *Ultramicroscopy* **63**, 227–238 (1996).
6. Carazo, J. M., Donate, L. E., Herranz, L., Secilla, J. P. & Carrascosa, J. L. Three-dimensional reconstruction of the bacteriophage $\Phi 29$ at 1.8 nm resolution. *J. Mol. Biol.* **192**, 853–867 (1986).
7. Knapek, E. Properties of organic specimens and their supports at 4 K under irradiation in an electron microscope. *Ultramicroscopy* **10**, 71–86 (1982).
8. Symons, M. C. R. The pre-knock-on concept. *Ultramicroscopy* **10**, 41–44 (1982).
9. Symons, M. C. R. Electron spin resonance studies of radiation damage to DNA and to proteins. *Radiat. Phys. Chem.* **45**, 837–845 (1995).

10. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).
11. Tadic, B. Nonuniversal scaling behavior of Barkhausen noise. *Phys. Rev. Lett.* **77**, 3843–3846 (1996).
12. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality. *Phys. Rev. A* **38**, 364–374 (1988).
13. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: an explanation of $1/f$ noise. *Phys. Rev. Lett.* **59**, 381–384 (1987).
14. Paczuski, M., Maslov, S. & Bak, P. Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E* **53**, 414–443 (1996).
15. Bragg, W. L. A Model illustrating intercrystalline boundaries and plastic flow in metals. *J. Sci. Instr.* **19**, 148–150 (1942).
16. Bragg, W. L. & Nye, J. F. A dynamical model of a crystal structure. *Proc. R. Soc. Lond. A* **190**, 474–481 (1947).
17. Massobrio, C., Pontikis, V. & Martin, G. Amorphization induced by chemical disorder in crystalline NiZr_2 : a molecular-dynamics study based on an n -body potential. *Phys. Rev. Lett.* **62**, 1142–1145 (1989).

Acknowledgements. The authors thank R. Schlögl and W. Storck for continuous support and fruitful discussions.

Correspondence and requests for materials should be addressed to R.S. (e-mail: schuster_r@thi-berlin.mpg.de).

Diamondoid hydrocarbons as indicators of natural oil cracking

J. E. Dahl*, J. M. Moldowan*, K. E. Peters†, G. E. Claypool†, M. A. Rooney†, G. E. Michael‡, M. R. Mello§ & M. L. Kohnen||

* Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305-2115, USA

† Mobil Exploration and Producing Technical Center, PO Box 650232, Dallas, Texas 75265-0232, USA

‡ Conoco Inc., PO Box 2197, Houston, Texas 77252-2197, USA

§ Research and Development Center, Petrobras, Cidade Universitaria, Quadra 7-Ilha do Fundao, 21949-900 Rio de Janeiro, Brazil

|| Shell International Exploration and Production, PO Box 60, 2280 AB Rijswijk, The Netherlands

Oil cracking—the thermal breakdown of heavy hydrocarbons to smaller ones—takes place within oil-bearing rock formations at depths commonly accessed by commercial oil wells. The process ultimately converts oil into gas and pyrobitumen, and thus limits the occurrence of petroleum and the success of exploration. Thermal cracking of liquid petroleum increases with depth until it reaches completion at the so-called ‘oil deadline’, which is generally placed^{1,2} at around 5 km depth and at temperatures of 150–175 °C. However, cracking experiments^{3–8} and the discovery of relatively ‘hot’ oil reservoirs^{9,10} imply that petroleum is thermally more stable than previously assumed; in fact it has been suggested that liquid petroleum might persist at temperatures reaching^{6,11–13} or even exceeding^{3,14} 200 °C. But reliable estimates of the extent of oil cracking and the depth at which it occurs in any given reservoir are difficult to obtain. Here we demonstrate that the relative abundance of diamondoids, a class of petroleum compounds whose unique thermal stability leads to their progressive concentration during cracking¹⁵, can be used to identify the occurrence and estimate the extent of oil destruction and the oil deadline in a particular basin. We are also able to identify oils consisting of mixtures of high- and low-maturity components, demonstrating that our method yields valuable information on the cracking and mixing processes affecting petroleum systems.

The occurrence of diamondoids in petroleum was first recognized in 1933, when adamantane, consisting of cyclohexane rings in ‘chair’ conformation (the diamond subunit), was isolated from Czechoslovakian crude oil¹⁶. Diamantane, consisting of two subunits, is the second pseudohomologue, triamantane the third, and so on. We believe that the diamondoids found in petroleum result from carbonium ion rearrangements of suitable organic precursors (such as multi-ringed terpene hydrocarbons) on clay mineral

superacids in the source rock during oil generation. The diamond structure endows these molecules with an unusually high thermal stability, relative to that of other crude-oil compounds, and allowed their isolation from petroleum by hydrocracking oil to gas¹⁵. Our method makes use of this diamondoid concentration effect to assess oil destruction in reservoirs due to either thermal cracking¹⁷ and/or thermochemical sulphate reduction (TSR)¹⁸. The latter process involves the oxidation of oil by sulphate at high temperatures within the reservoir.

Our methodology is conceptually analogous to determining the extent of evaporation of salt water by analysing salinity. For example, for a group of salt ponds derived from a common source (such as the Mediterranean Sea), it is possible to estimate the extent of evaporation by comparing the salinity of the ponds. A doubling of salinity implies 50% evaporation. Similarly, for a group of oils

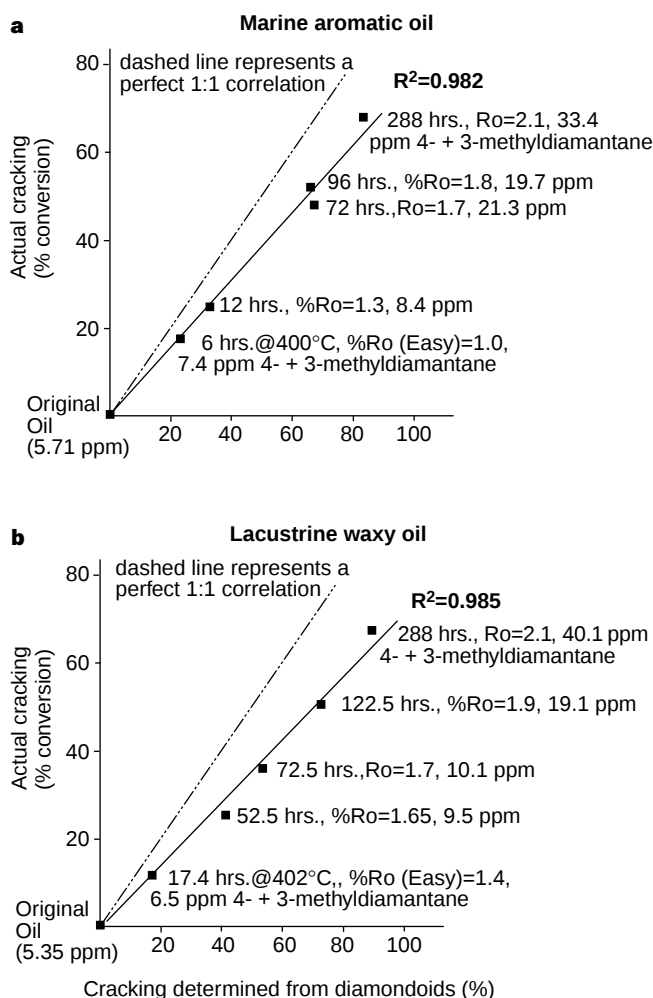


Figure 1 Correlation between extent of cracking and methyladamantane concentration in laboratory oil-cracking experiments. Experiments were done in evacuated vessels, anhydrous, at 400 °C and 402 °C for a marine source rock-derived aromatic oil from Prudhoe Bay, Alaska (**a**), and a lacustrine waxy oil from Indonesia (**b**), respectively. Actual cracking percentages were determined by mass balance by comparing the original to the final weight of the oil. The correlation between actual cracking from mass balance and that determined from diamondoid concentrations according to the equation given in the text is excellent (R^2 for both is greater than 0.98). However, the diamondoid method consistently overpredicts the extent of cracking, especially in the highly cracked samples. This discrepancy is believed to be due primarily to the evaporative loss of light ends in the oils before diamondoid analysis (samples were stored and shipped before analysis). The vitrinite reflectance equivalence of the oils was calculated using the ‘Easy%Ro’ method of Sweeney and Burnham²⁷.

from the same source (such as the Smackover Formation, Gulf of Mexico), the extent of oil destruction due to cracking can be estimated by comparing diamondoid concentrations in cracked and uncracked oils.

To illustrate this method, we performed a series of cracking experiments in the laboratory. In these experiments, we monitored the concentration of methyladamantanes, C_{15} diamondoids that are not readily lost to evaporation. The original oils were of two different types, a waxy, lacustrine oil from Indonesia, and an aromatic, marine oil from Alaska. Our results (Fig. 1) show that the increase in methyladamantane concentration is directly proportional to the extent of cracking (the percentage of liquid converted to gas and pyrobitumen), indicating that under the conditions of our experiments, diamondoids are neither destroyed nor created. Instead, they are conserved and concentrated, and hence can be considered a naturally occurring "internal standard" by which the extent of cracking can be determined. The extent of cracking (the percentage conversion of liquid hydrocarbons to gas and pyrobitumen) is equal to $[1 - (C_o/C_c)] \times 100$, where C_o is the concentration of methyladamantanes in the uncracked samples (here termed

"diamondoid baseline") and C_c is the methyladamantane concentration of any cracked sample derived from the same starting oil.

In nature, it is not possible to analyse the concentrations of diamondoids in an oil before cracking, so we must infer the diamondoid baseline by analysing a group of uncracked, non-biodegraded and non-fractionated oils from the same source. This is in fact not difficult, as there are published methodologies used routinely by geochemists to recognize low-maturity, non-biodegraded and non-fractionated oils^{1,19,20}. However, each source has its own baseline, which must be determined individually. For example, the diamondoid baseline for oils from the Monterey Formation of California is between 1 and 2 p.p.m. This was determined by analysing 12 non-biodegraded, non-fractionated Monterey oils with a range of thermal maturities (based on biomarker ratios and API gravities ranging from ~15 to ~35). Based on analysis of 10 oils ranging in colour from black to pale yellow, the baseline for Austin Chalk oils of south Texas is between 7 and 10 p.p.m. The diamondoid baseline for oils sourced from the Barreirinhas Member of the Curua Formation, Upper Devonian, Solimoes basin, Brazil, is in the 2–5 p.p.m. range.

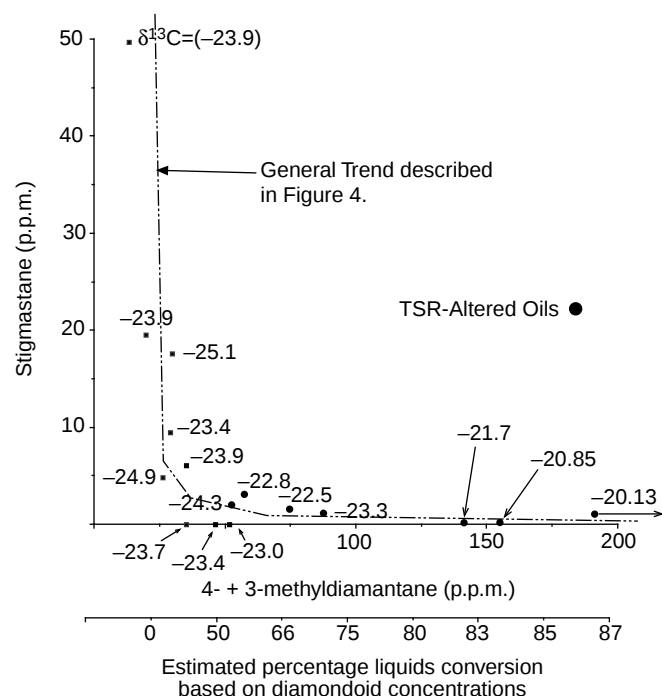


Figure 2 Correlation between biomarker (stigmastane) and diamondoid (methyladamantane) concentration in a series of Smackover-sourced oils from southwestern Alabama. Whole-oil carbon-isotope signatures ($\delta^{13}C$) are plotted next to the data points and become more positive as methyladamantane concentrations increase. These oils have undergone varying degrees of thermal cracking and/or thermochemical sulphate reduction (TSR), processes which generate isotopically light gases, and as a result the residual oil has become isotopically heavy¹⁷. Through the process of oil destruction due to cracking and TSR, the diamondoids have been dramatically concentrated (by a factor of 5 or greater in some cases). Methyladamantane concentrations were determined by spiking the samples with a known amount of 1, d_3 -methyladamantane, followed by several HPLC and zeolite concentration procedures. Samples were then analysed by (MRM)-GC-MS (metastable reaction monitoring) monitoring the m/z 202⁺ → 187⁺ and the m/z 205⁺ → 187⁺ transitions for methyladamantanes and 1, d_3 -methyladamantane respectively. Stigmastane concentrations were determined from the metastable m/z 400⁺ → 217⁺ ion transition using cholane (m/z 330⁺ → 217⁺) as an internal standard, and an oil of known biomarker concentration also spiked with cholane as an external standard²³.

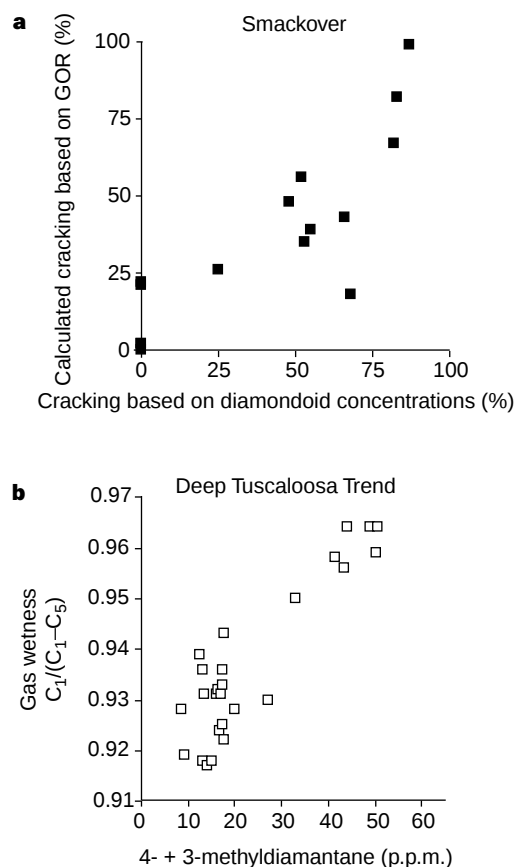


Figure 3 Comparison between different indicators for the extent of oil cracking. **a**, Correlation between "% conversion to gas", calculated by Claypool and Mancini¹⁷ from gas/oil ratios (GORs; see their Table 2) for the Smackover oils and the corresponding values derived from our diamondoid measurements. Given the number of factors that determine GOR, there is reasonably good agreement between the estimated extent of cracking based on diamondoids (shown on the x-axis) and that calculated by Claypool and Mancini¹⁷ from GOR (shown on the y-axis). **b**, Gas wetness, $C_1/(C_1-C_5)$, which is related to GOR, versus diamondoid concentrations for a series of condensates from the Deep Tuscaloosa Trend in southern Louisiana. These condensates are sourced from, and reservoir in, the Tuscaloosa Formation. Reservoir temperatures range from 165 to 195 °C at depths between 4.9 and 6.2 km. The correlation between gas wetness and diamondoid concentration suggests that gas wetness for these condensates is largely controlled by oil cracking.

One of the most straightforward examples of the increase in diamondoid concentrations due to natural oil cracking occurs in the Solimoes basin, Brazil. It has been previously recognized that oil in some reservoirs has been extensively cracked due to intense heat from igneous (diabase) intrusions²¹. Six uncracked oils in reservoirs distant from any intrusion (more than 850 m) were analysed to determine a diamondoid baseline, which for Barreirinhas-sourced oil is between 2 and 5 p.p.m. Oil in the SOJ-1AM reservoir, which occurs within 150 m of an intrusion, has been extensively cracked²¹. Vitrinite reflectance in surrounding rocks has been increased to 1.8% and the methyl-diamantanes in the oil have been dramatically concentrated (just as in the pyrolysis experiments) to 32 p.p.m. Comparing the estimated original methyl-diamantane concentration in this sample (2–5 p.p.m.) with the measured value of 32 p.p.m., we estimate (on the basis of the equation for extent of cracking mentioned above) that 85–94% of the original liquids have been converted to gas and pyrobitumen.

Other examples of natural diamondoid concentration due to thermal cracking and TSR reactions occur in the Norphlet Trend in the Gulf of Mexico; for example, the Mary Ann and 823 fields in Mobile Bay¹⁷. Here liquid petroleum may be almost entirely converted to gas as temperatures near 200 °C. The highly resistant diamondoids, however, survive, and can become so concentrated in the liquid condensate that they crystallize out during production, causing plugging problems in wells.

The destruction of oil due to cracking and TSR in the Norphlet can be observed from the carbon-isotope signature of the oil or condensate. The oils we analysed were generated from the Smackover Formation. The uncracked, non-TSR-altered oils had isotopic signatures between –23.4‰ and –25.1‰. However, as cracking and TSR generate isotopically light gases, the remaining liquids may become isotopically heavy. As a result, the most cracked and TSR-

altered Smackover-sourced oils studied here range from –21.7‰ to –20.13‰. These same samples have highly elevated diamondoid concentrations, over 5 times that of the uncracked and non-TSR-altered samples (Fig. 2). Based on this study, cracking of Smackover oils appears to begin (diamondoid concentrations begin to rise) at reservoir temperatures around 160 °C for non-TSR-altered oils. A rise in diamondoid concentrations suggests that TSR begins oil destruction at about 120–130 °C.

Oil destruction due to cracking can also be estimated from gas/oil ratios (GORs)¹⁷, but GORs reflect not only oil cracking, but also the GOR ratios of the primary product of petroleum formation, and are affected by fractionation and diffusion processes within reservoirs and during petroleum migration by gas flushing of reservoirs (displacement of oil by gas). Although the use of GORs as an indicator of oil cracking can thus make for tenuous arguments, we observe reasonable agreement between the “% conversion to gas” calculated by Claypool and Mancini¹⁷ from Smackover oil GOR values and the extent of cracking inferred from our diamondoid measurement (Fig. 3a). We have also found a fairly good correlation between gas wetness, which is related to GOR, and diamondoid concentrations for a series of condensates from the Deep Tuscaloosa Trend in Louisiana (Fig. 3b), which suggests that the gas wetness for these condensates is largely controlled by oil cracking.

In addition to diamondoid concentrations, it is informative to determine biomarker concentrations in the same oils and condensates. Biomarkers are molecular fossils which have the recognizable carbon structures of natural products synthesized by oil precursor organisms such as bacteria and algae^{22,23}. For example, cholestane, derived from cholesterol, can be found in almost every oil, even those over 500 million years old²⁴. Unlike diamondoids, they generally contain strained bonds that can be cracked easily compared to other oil components. Therefore, they decrease in concentration with thermal maturity²⁵ before cracking of the major oil components (Fig. 4). Different biomarkers vary greatly in terms of thermal stability. We prefer using the concentration of one of the less stable biomarkers, 5 α ,14 α ,17 α (H)-24-ethylcholestane 20R (stigmastane, a C₂₉ sterane derived from algae and higher plants) because its concentration drops to near zero at about the point where diamondoid concentrations begin to rise (Figs 2 and 4). Plotting both biomarkers and diamondoids allows us to recognize “mixed” oils derived from both a low-maturity, biomarker-rich source and a high-maturity, highly cracked, diamondoid-rich source (Fig. 4).

It is important to remember when using this method that factors other than oil cracking can affect concentrations of both diamondoids and biomarkers. Two examples are biodegradation and reservoir fractionation. However, these factors can usually be recognized²⁶, and after they have been taken into account, it is then possible to use diamondoid–biomarker curves along with geological information to identify and characterize the particular basin. The information provided by our approach will be particularly valuable for computer basin models used to predict oil cracking in a particular basin and improve our understanding of regional gas and oil distributions. □

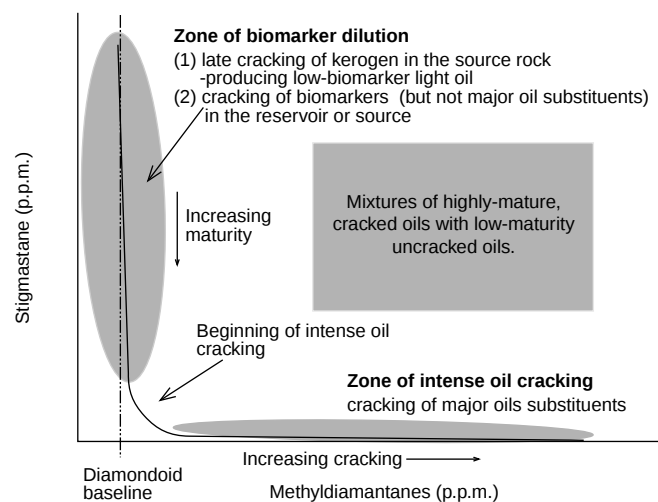


Figure 4 Illustration of the asymptotic curve characterizing the correlation between the concentrations of diamondoids (methyl-diamantanes) and biomarkers (stigmastane) in oils of different thermal maturities, cracked and uncracked, from a specific petroleum system. Although the shape for each generative petroleum system is similar, the original diamondoid and biomarker concentrations may be quite different. The shape of the curve is the result of decreasing biomarker concentrations with increasing thermal maturity (moving down the y axis), followed by oil cracking, resulting in increased diamondoid concentrations due to their high thermal stability. Stigmastane 20R was chosen for the y-axis because it disappears at about the time diamondoid concentrations begin to rise. The use of rearranged steranes (diasteranes), which are more thermally stable, results in a much flatter curve (that is, closer to a straight line), indicating that stigmastane is almost totally cracked before the major constituents in the oil, whereas the more stable diasteranes are cracked along with the major constituents in the oil.

Received 8 April 1998; accepted 18 February 1999.

- Hunt, J. *Petroleum Geochemistry and Geology* (Freeman, San Francisco, 1996).
- Landes, K. K. Eometamorphism and oil and gas in time and space. *Am. Assoc. Petrol. Geol. Bull.* **51**, 828–841 (1967).
- Mango, F. D. The stability of hydrocarbons under the time-temperature conditions of petroleum genesis. *Nature* **352**, 146–148 (1991).
- Burnham, A. et al. Decomposition kinetics and mechanism of n-hexadecan-1,2-¹³C₂ and dodec-1-ene-1,2-¹³C₂ doped in petroleum and n-hexadecane. *Geochim. Cosmochim. Acta* **61**, 3725–3737 (1997).
- Horsfield, B., Schenk, H. J., Mills, N. & Welte, D. H. in *Advances in Organic Geochemistry 1991* (eds Eckardt, C. B., Maxwell, J. R., Larter, S. R. & Manning, D. A. C.) *Org. Geochem.* **19**, 191–204 (1992).
- Behar, F. & Vandenbroucke, M. Experimental determination of the rate constants of the n-C₂₅ thermal cracking at 120, 400, and 800 bar: Implications for high-pressure/high temperature prospects. *Energy Fuels* **10**, 932–940 (1996).
- Jackson, K. J., Burnham, A. K., Braun, R. L. & Knauss, K. G. Measurement of pressure effects on n-hexadecane cracking rates. *Org. Geochem.* **23**, 941–953 (1995).
- Schenck, H. J., Di Primio, R. & Horsfield, B. The conversion of oil into gas in petroleum reservoirs. Part 1: Comparative kinetic investigation of gas generation from crude oil of lacustrine, marine and

- fluviodeltaic origin by programmed-temperature closed-system pyrolysis. *Org. Geochem.* **26**, 467–481 (1997).
9. Pepper, A. S. & Dodd, T. A. Simple kinetic models of petroleum formation. Part II: Oil-gas cracking. *Mar. Petrol. Geol.* **12**, 321–340 (1995).
 10. Price, L. C. Organic geochemistry of core samples from an ultra-deep hot well (300 °C, 7 km). *Chem. Geol.* **37**, 215–228 (1982).
 11. Quigley, T. M. & McKenzie, A. S. The temperature of oil and gas formation in the subsurface. *Nature* **333**, 549–552 (1988).
 12. Braun, R. L. & Burnham, A. K. PMOD: A flexible model of oil and gas generation, cracking and expulsion. *Org. Geochem.* **19**, 161–172 (1992).
 13. Ungerer, P., Behar, F., Billalba, M., Heum, O. R. & Audibert, A. in *Advances in Organic Geochemistry 1987* (eds Mattavelli, L. & Novelli, L.) *Org. Geochem.* **13**, 857–868 (1988).
 14. Price, L. C. Thermal stability of hydrocarbon in nature: Limits, evidence, characteristics and possible controls. *Geochim. Cosmochim. Acta* **57**, 3261–3280 (1993).
 15. Petrov, A. A., Arefjev, O. A. & Yakubson, Z. V. in *Advances in Organic Geochemistry 1973* (eds Tissot, B. & Biener, F.) 517–522 (Editions Technip, Paris, 1974).
 16. Landa, S., Machacek, V., Mzourek & Landa *Chim. Ind.* (special no. 506) (1933) (in French).
 17. Claypool, G. E. & Mancini, E. A. Geochemical relationships of petroleum in Mesozoic reservoirs to carbonate source rocks of Jurassic Smackover Formation, Southwestern Alabama. *Am. Assoc. Petrol. Geol. Bull.* **73**, 904–924 (1989).
 18. Orr, W. Changes in sulfur content and isotopic ratios of sulfur during petroleum maturation—Study of Big Horn Basin Paleozoic oils. *Am. Assoc. Petrol. Geol. Bull.* **58**, 2295–2318 (1974).
 19. Schleyer, P. A simple preparation of adamantane. *Am. Chem. Soc.* **79**, 3292 (1957).
 20. Mello, M. R., Gonçalves, F. T., Babinsky, N. A. & Miranda, F. P. Hydrocarbon migration and its near-surface expression. *Mem. Am. Assoc. Petrol. Geol.* **66**, 401–411 (1996).
 21. Eglinton, G., Scott, P. M., Besky, T., Burlingame, A. L. & Calvin, M. Hydrocarbons of biological origin from a one-billion-year-old sediment. *Science* **145**, 263–264 (1964).
 22. Eglinton, G. & Calvin, M. Chemical fossils. *Sci. Am.* **261**, 32–43 (1967).
 23. Peters, K. E. & Moldowan, J. M. *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments* (Prentice Hall, Englewood Cliffs, NJ, 1993).
 24. Summons, R. E., Powell, T. G. & Boreham, C. J. Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, Northern Australia: III. Composition of extractable hydrocarbons. *Geochim. Cosmochim. Acta* **52**, 1747–1763 (1988).
 25. Requejo, A. J. in *Biological Markers in Sediments and Petroleum* (eds Moldowan, J. M., Albrecht, P. & Philp, R. P.) 223–240 (Prentice Hall, Englewood Cliffs, NJ, 1992).
 26. Thompson, K. F. M. Classification and thermal history of petroleum based on light hydrocarbons. *Geochim. Cosmochim. Acta* **47**, 303–316 (1983).
 27. Sweeney, J. J. & Burnham, A. K. Evaluation of a simple model of vitrinite reflectance based on chemical kinetics. *Am. Assoc. Petrol. Geol. Bull.* **74**, 1559–1570 (1990).

Correspondence and requests for materials should be addressed to J.E.D. (e-mail: Dahl@Pangea.Stanford.edu).

Early hominid stone tool production and technical skill 2.34 Myr ago in West Turkana, Kenya

H. Roche*, A. Delagnes*, J.-P. Brugal†, C. Feibel‡, M. Kibunjia§, V. Mourre|| & P.-J. Texier||

* CNRS, Préhistoire et Technologie, MAE (boite 3), 21 allée de l'Université, 92023 Nanterre Cedex, France

† CNRS, UMR 6636, MMSH, 5 rue du Château de l'Horloge, 13094 Aix en Provence, France

‡ Department of Anthropology, Rutgers University, New Brunswick, New Jersey 08903, USA

§ National Museums of Kenya, PO Box 40658, Nairobi, Kenya

|| CNRS, Préhistoire et Technologie, CRA, Sophia Antipolis, 06565 Valbonne Cedex, France

Well-documented Pliocene archaeological sites are exceptional. At present they are known only in East Africa, in the Hadar^{1,2} and Shungura³ formations of Ethiopia and in the Nachukui formation of Kenya. Intensive archeological survey and a series of test excavations conducted in the Nachukui formation since 1987 have led to the discovery of more than 25 archaeological sites whose ages range from 2.34 to 0.7 million years before present (Myr)^{4,5}, and to the extensive excavation of two 2.34-Myr sites, Lokalelei 1 in 1991 (refs 6, 7) and Lokalelei 2C in 1997. Lokalelei 2C yielded nearly 3,000 archaeological finds from a context of such good preservation that it was possible to reconstitute more than 60 sets of complementary matching stone artefacts. These refits, predating the Koobi Fora refits by 500 Kyr (ref. 8), are the

oldest ever studied. Here we describe a technological analysis of the core reduction sequences, based on these refits, which allows unprecedented accuracy in the understanding of flake production processes. We can thus demonstrate greater cognitive capacity and motor skill than previously assumed for early hominids, and highlight the diversity of Pliocene technical behaviour.

Lokalelei 2C (LA2C; 3° 56' 57.0" N, 35° 46' 43.7" E) is situated in badlands exposures within the drainage of the Lokalelei sand bed ephemeral stream. The exposed sequence includes strata of the Lokalelei and Kalochoro members of the Nachukui formation⁹ (Fig. 1). Near the site, both the Kokiselei and Ekalalei tuffs are exposed, and the archaeological site is 9 m above the latter (Fig. 2). Lokalelei 1 (ref. 9) is about 1 km away, at approximately the same stratigraphic level. A distinctive mollusc-packed sandstone 12 m below the site is used for local correlation.

The Kokiselei tuff has been correlated with Tuff E of the Shungura formation⁹, with a stratigraphically scaled age of 2.40 ± 0.05 Myr (ref. 10). The Ekalalei tuff is a geochemical correlate of Shungura tuff F-1 (ref. 9). This tuff lies immediately above Tuff F in the Omo (termed the Kalochoro tuff in West Turkana), which has been isotopically dated to 2.34 ± 0.04 Myr (ref. 10). The Ekalalei tuff (F-1) is thus slightly younger than 2.34 Myr, and is further constrained by an age of 2.32 Myr on Tuff G¹¹. In the Shungura sections, Tuff G lies 30–45 m above Tuff F, and thus we can reasonably estimate an age of close to 2.34 ± 0.05 Myr for the Lokalelei archaeological sites.

The depositional sequence represented in the Lokalelei exposures includes a succession of fluvial facies separated by lacustrine deposits (Fig. 2), which probably correlate with the Lokeridede lake phase of the Turkana basin¹². The Lokalelei sites are within floodplain paleosols in the upper fluvial strata, reflecting low energy, but proximal rather than distal, environments. LA2C is stratified in a clayey sand lens within a floodplain vertisol. The palaeogeographic setting of the site places it near the intersection of ephemeral basin-margin streams and the meandering, axial, ancestral Omo river.

A total of 17 m² was excavated at LA2C. The *in situ* material, comprising abundant artefacts and fewer faunal remains, formed a dense concentration (mean, 129 per m²), with a very limited horizontal extension (10 m²). It is vertically dispersed by up to 50 cm, largely by pedogenic churning of the associated vertic paleosols. Despite this dispersion, the assemblage has remarkable archaeological homogeneity and preservation, which are demonstrated by the very fresh condition of the artefacts, the high proportion (28%) of small elements (<1 cm), the exceptional percentage of the refitted pieces (10% of the whole assemblage and 20% of the *in situ* material have been refitted) and the spatial distribution of refits, which corresponds closely with the distribution of the whole assemblage (Fig. 3).

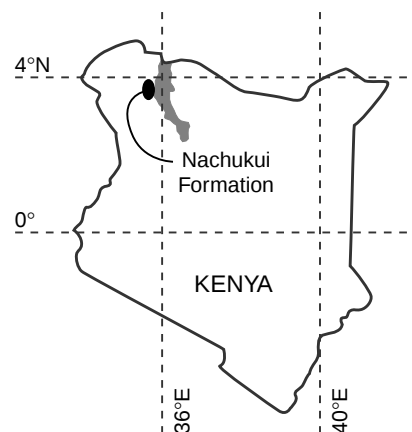


Figure 1 Map showing the location of the Nachukui formation.

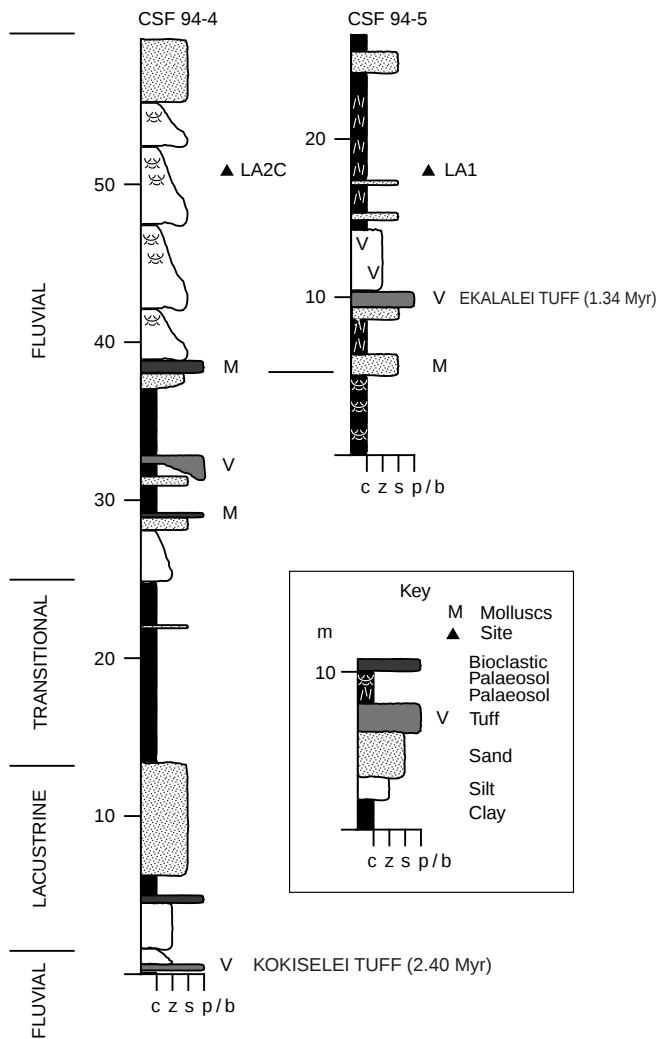


Figure 2 Stratigraphic sections through the Lokalelei archaeological sites. The sections represent the upper Lokalelei member and lower Kalochoro member, with the member boundary placed at about the level of the mollusc-packed

sandstone, by which the two sections shown here are correlated. c, clay; z, silt; s, sand; p/b, pyroclastic or bioclastic sediment.

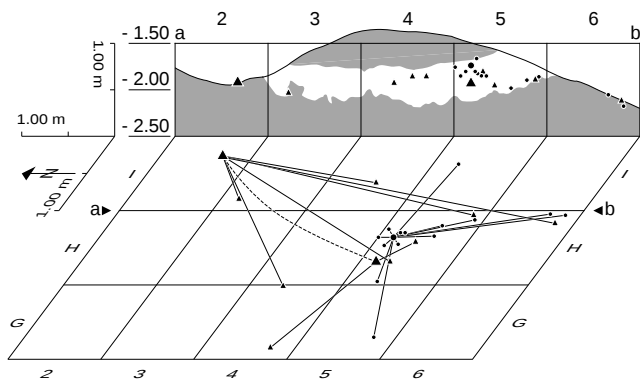


Figure 3 Horizontal and vertical distribution of R35 (triangles) and R9 (circles) refits. Three unplotted surface elements from R35 refit are not mapped. On the horizontal view, lines connect the flakes to their core. The dotted line joins two cores belonging to R35 refit, which were separately flaked after the block was broken. The arrowheads marked a and b indicate the horizontal axis along which the related vertical section was taken. G, H, I: west-east notation of grid squares; Numbers: north-south notation of grid squares.

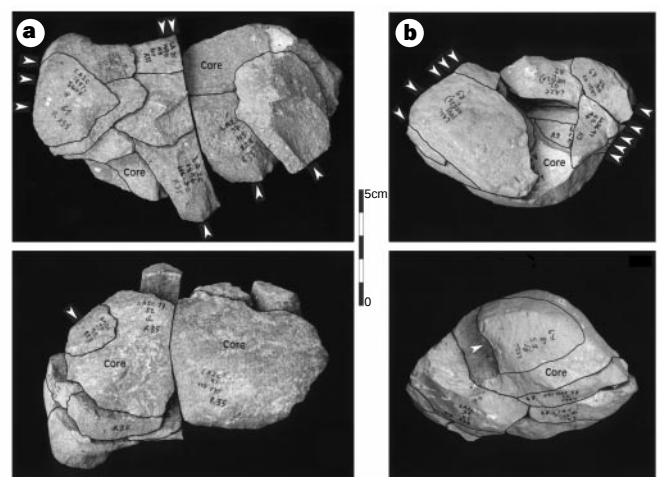


Figure 4 R35 and R9 refits. **a**, R35 refit (2 cores and 11 flakes). **b**, R9 refit (1 core and 14 flakes). These two refits are characteristic of the main reduction sequence: unidirectional or multidirectional removals flaked on a single debitage surface, from natural or prepared platforms. Arrows indicate the direction of the removals.

Table 1 Faunal remains (NISP) from LA2C

	Excavation	Surface	Total
Lamelliibranches	1	1	2
Gasteropods	1		1
Fish (Siluriformes)	1	1	2
Chelonians	19	8	27
Crocodylidae (<i>Crocodylus</i> sp.)	5	3	8
Reptile indet.	42	2	44
Aves (<i>Sthrutio</i> sp.)	17		17
<i>Hystrix</i> sp.		1	1
Bovid indet.	10		10
Bovid small	5	2	7
Bovid medium	3	3	6
Antilopini aff. <i>Gazella</i> sp.	2		2
Reduncini medium		2	2
Alcelaphini small	1	1	2
Hippotragini		1	1
Suid indet.		3	3
<i>Kolpochoerus limnetes</i>		1	1
Equid indet.		3	3
<i>Hipparion</i> aff. <i>ethiopicum</i>	2	1	3
Hippopotamid indet.		2	2
<i>Ceratotherium</i> sp.	9	1	10
Fragment indet.	99	109	208
Mammal large	2	4	6
Mammal indet.	22		22
Total	241	149	390

Faunal remains from LA2C represent 12 mammal species as well as various reptiles and fish. The excavation and surface zone yield, respectively, 241 and 149 faunal specimens (Table 1). Most of the material found *in situ* consisted of reptile bones, often fractured by compaction or pedogenesis. Shell fragments of a large land tortoise, together with several fragments of ostrich egg shell, are the best preserved elements of this relatively small faunal assemblage. Mammals are represented mainly by teeth and include bovids, suids, equids and a hyposodont rhinocerotid, attributed to *Ceratotherium simum germanoaffricanum*. With the addition of this rhinocerotid and a large rodent (*Hystrix* sp.), the mammals from LA2C are characteristic of the species already recorded in the Kalocho member⁹. They are essentially grazing species, indicating an open environment on an alluvial plain, with patches of bushes or forest along an ephemeral river. The fragmentation and the distribution of the fossils within the presumably short sequence of LA2C point to natural accumulation. The bones, which are encrusted and poorly preserved, show no evidence of anthropic action. However, the tortoise bones and ostrich-egg fragments are more closely associated with the lithic artefacts; their systematic presence in both Lokalei sites may show a possible hominid collecting strategy.

The lithic assemblage from LA2C (Table 2) is dominated by debitage products (91%). These elements, associated with the cores and hammerstones, indicate that complete sequences (*chaînes*

opératoires) of lithic production have taken place at the site. The rest of the assemblage consists of worked and unworked cobbles. The raw materials, mainly lavas (basalt and phonolite), were gathered from the ephemeral streams draining the Murua Rith range, which borders the Turkana basin on the west. Ten different categories have been identified, probably reflecting the diversity of raw materials available in the streams.

Study of the refitted material indicates that 60–70 cobbles were knapped on the spot to produce flakes (Table 2). Very few of these flakes were modified by subsequent retouching. Cores range from coarse-grained cobbles, from which only a few flakes have been removed, to cores of fine-grained lava pebbles, from which a large series of flakes (up to 30) have been produced. At LA2C, the dominant reduction sequence consists of these more reduced cores. Unidirectional or multidirectional removals are flaked on a single debitage surface, from natural or prepared platforms (Fig. 4). This knapping system allows the production of a long series of removals without changing the volumetric structure of the core. The repeated application by the knappers of the same technical principles to a whole series of cores, and during the reduction of each core, indicates an elaborate debitage scheme, implying motor precision and coordination. These principles include an appreciation of the quality of the collected raw materials, a judicious exploitation of the natural morphology of the blocks and the maintenance of adequate flaking angles during the entire debitage sequence. These show that the notion of production was already assimilated by a group of hominids in this particular area. This notion is integrated within a real debitage strategy, here well-mastered and unprecedented for this period.

Overall simplicity and similarities between assemblages are the two main arguments recently put forward to substantiate a technological stasis hypothesis between the 2.6 and 1.6 Myr time periods², and to merge the related assemblages into a single vast 'Oldowan' technocomplex. The stasis hypothesis cannot hold out against the detailed technological analysis of the LA2C lithic assemblage. There can be no doubt about the elaborate character of the LA2C debitage schemes, which are far more sophisticated than at any other Pliocene site. The assemblage also provides yet another example of the technical diversity within this time span. It should be stressed, however, that temporal variation is not necessarily in the direction of ever greater sophistication. Given the so-far largely underestimated spatio-temporal distance separating Plio-pleistocene 'Oldowan' assemblages¹³, the variation observed probably reflects technical solutions to different environments and needs, as well as differences in cognitive and motor skills among early hominid groups characterized by non-synchronous evolutionary processes. □

Received 21 December 1998; accepted 19 March 1999.

Table 2 Composition of the lithic assemblage from LA2C

Category	Excavation	Surface	Total	Per cent of assemblage	Refitted	Per cent of category
Whole flakes	349	143	492	19.0	94	19.1
Broken flakes	491	260	751	29.1	95	12.6
Small flakes (< cm) and chips	690	27	717	27.8		
Fragments indet.	332	62	394	15.3	37	9.4
Retouched flakes	9	4	13	0.5	4	31
Whole cores	51	11	62	2.4	24	30.8
Broken cores	9	1	10	0.4	6	60.0
Chopper cores	1		1			
Tool cores	2	4	6			
Hammerstones	17		17	0.7		
Worked pebbles	16	2	18	0.7		
Broken pebbles	53	2	55	2.1	4	7.3
Unmodified pebbles	47		47	1.8		
Total	2,067	516	2,583	100.0	264	10.2

Of the 63 sets of refits which have been reconstituted, 50 sets comprise between 2 and 5 artefacts (51.8% of the refitted material), 9 sets between 6 and 10 artefacts, and the last 4 sets 11, 13, 15 and 20 artefacts, respectively. In 24 sets, the core is included. Some cobbles are totally or almost totally reconstructed (Fig. 4b). Other sets contain a large number of elements (for example, R33 refit, 20 elements) but are far from complete. Fragments indet.: non-orientable fragments.

1. Roche, H. & Tiercelin, J.-J. in *Proc. VIIth Panafrican Congress of Prehistory and Quaternary Studies* (eds Leakey, R. E. F. & Ogott, B.) 194–199 (Tilmiap, Nairobi, 1980).
2. Semaw, S. *et al.* 2.5-million-year-old stone tool from Gona, Ethiopia. *Nature* **385**, 333–336 (1997).
3. Howell, F. C., Heesaerts, P. & de Heinzelin, P. Depositional environment, archaeological occurrences and hominids from members E and F of Shungura Formation (Omo Basin, Ethiopia). *J. Hum. Evol.* **16**, 643–664 (1987).
4. Kibunjia, M., Roche, H., Brown, F. H. & Leakey, R. E. Pliocene and Pleistocene archaeological sites west of Lake Turkana, Kenya. *J. Hum. Evol.* **23**, 431–438 (1992).
5. Roche, H. & Kibunjia, M. Les sites archéologiques plio-pléistocènes de la Formation de Nachukui, West Turkana, Kenya. *Cr. Acad. Sc. Paris* **318**, 1145–1151 (1994).
6. Kibunjia, M. Pliocene archaeological occurrences in the lake Turkana Basin, Kenya. *J. Hum. Evol.* **27**, 157–171 (1994).
7. Kibunjia, M. *Archaeological Investigations of Lokalalei 1 (Gajh 5): A Late Pliocene Site, West of Lake Turkana, Kenya*. Thesis, Univ. Rutgers (1998).
8. Isaac, G. L. (ed.) *Koobi Fora Research Project Vol. 5* (Clarendon, Oxford, 1997).
9. Harris, J. M., Brown, F. H. & Leakey, M. G. Stratigraphy and paleontology of Pliocene and Pleistocene localities west of Lake Turkana, Kenya. *Contrib. Sci.* **399**, 1–128 (1988).
10. Feibel, C. S., Brown, F. H. & McDougall, I. Stratigraphic context of fossils hominids from the Omo Group deposits, Northern Turkana Basin, Kenya and Ethiopia. *Am. J. Phys. Anthropol.* **78**, 595–662 (1989).
11. Brown, F. H., McDougall, I., Davis, T. & Maier, R. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 82–90 (Alan R. Liss, New York, 1985).
12. Feibel, C. S., Harris, J. M. & Brown, F. H. in *Koobi Fora Research Project Vol. 3* (ed. Harris, J. M.) 321–370 (Clarendon, Oxford, 1991).
13. Roche, H. in *XIII^e Int. Congress of Prehistoric and Protohistoric Sciences, Colloquia Vol. 4*, 55–63 (ABACO, Forlì, 1996).

Acknowledgements. We thank the Government of Kenya for permission to carry out this research and the National Museums of Kenya, Meave Leakey and the Kalokol Project for logistical support. The field work was funded by the French Foreign Affairs Ministry (Sous-Direction des Sciences Sociales, Humaines et de l'Archéologie). We thank Total (Kenya) for vehicles and fuel donation, and Crédit Agricole-Indosuez Bank for financial support. The team included J. Ekiyeni, T. Harkaker, B. K. Kimeu, B. Kimolo, R. M. Kioko, B. Kyongo, E. Lokokodi, S. Mekwea, V. Mugao, K. S. Mulwa, D. Mutinda, A. Mutisya, S. Mutuku, A. Nares, L. K. Ngolokerem, T. Numbé and M. Twaha. We thank J. Feblot-Augustins for her help in translating and editing this paper.

Correspondence and requests for materials should be addressed to H.R. (e-mail: roche@mae.u-paris10.fr).

The wing of *Archaeopteryx* as a primary thrust generator

Phillip Burgers* & Luis M. Chiappe†

* San Diego Natural History Museum, PO Box 121390, San Diego, California 92112, USA

† Vertebrate Paleontology, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, California 90007, USA

Since the late 1800s, the debate on the origin of flight in birds has centred around two antagonistic theories: the arboreal (take-off from trees) and cursorial (take-off from running) models^{1–6}. Despite broad acceptance of the idea that birds evolved from bipedal and predominantly terrestrial maniraptoriform dinosaurs^{1,7}, the cursorial model of flight origins has been less successful than the arboreal model. Three issues have contributed to this lack of success: the gap between the estimated maximum running speed of *Archaeopteryx* (2 metres per second) and its estimated minimum flying speed (6 metres per second)⁸; the high energy demands of evolving flight against gravity^{2,3}; and the problem of explaining the origin of the 'flight' stroke in an earthbound organism^{3,4}. Here we analyse the take-off run of *Archaeopteryx* through lift-off from an aerodynamic perspective, and emphasize the importance of combining functional and aerodynamic considerations with those of phylogeny^{1,9,10}. Our calculations provide a solution to the 'velocity gap' problem and shed light on how a running *Archaeopteryx* (or its cursorial maniraptoriform ancestors) could have achieved the velocity necessary to become airborne by flapping feathered wings.

Although, as a flier, it probably represents a relatively late stage in the evolution of bird flight, *Archaeopteryx* plays a central role in the debates on the origins of flight^{2–6,11}. Proponents of the arboreal model consider *Archaeopteryx* to have been a tree climber, but evidence in support of this is weak at best^{1,12–14}. Despite lacking the pulley-like action of the supracoracoid muscle of modern birds, which probably limited its capacity for carrying out fast, high-

amplitude wing beats¹⁵, *Archaeopteryx*'s pectoral musculature was apparently sufficient for flapping¹⁶. This conclusion is also supported by the lateral orientation of its glenoid facet¹⁷ and the passive pronation-supination of its hand, as inferred from its wrist morphology¹⁸. Thus, *Archaeopteryx* appears to have been a predominantly terrestrial animal that, given the limited volume of its pectoral muscles and the relatively low amplitude of its wing beat, presumably had to run to take off^{9,19}, flapping its wings in a fashion similar to that of large extant birds²⁰.

Our aerodynamic model begins with *Archaeopteryx* initiating the take-off run with forward propulsion generated by its hindlimbs at the same time as it starts flapping its wings (Figs 1, 2). Calculations indicate that, during the take-off run, the initial hindlimb-supplied propulsion is gradually replaced by wing thrust (see Methods). Simultaneously, the lift generated by the wings—here called 'residual' as it does not exert work on the bird until lift-off—'unloads' the hindlimbs of the body weight (Fig. 2). This dual force migration (propulsion and body weight support) from the hindlimbs toward the wings has profound implications for the estimated maximum running speed of *Archaeopteryx*. Clearly, flapping increases the bird's running speed. As the residual lift due to flapping relieves the hindlimbs of body weight support, its running speed is further increased, which, in turn, increases the residual lift (which increases with the square of the running speed). At a certain point in the take-off run, the residual lift becomes greater than the bird's weight and so is converted to useful lift: *Archaeopteryx* takes off. At this point, lift becomes a force that exerts work on the bird. Wing thrust is now the sole source for generating the velocity necessary for sustained lift.

Previous calculations for the maximum running speed of *Archaeopteryx* assumed that its hindlimbs alone generated propulsive force and provided support for its full weight during the take-off run⁸. However, when the proposed upward force migrations are considered, *Archaeopteryx* can reach its estimated minimum flying speed (6 m s⁻¹ in ref. 8), 7.8 m s⁻¹ in our model) by means of the thrust and residual lift produced by its flapping wings. Our calculations indicate that, 3 s after beginning its take-off run, *Archaeopteryx* would have achieved a speed of 7.8 m s⁻¹. Extant lizards are known to have burst speeds which last for much longer times²¹, and there is no indication that *Archaeopteryx* was metabolically incapable of the same. Thus, the 'velocity gap' ceases to exist.

This study indicates that *Archaeopteryx*'s wings may have been an efficient aerodynamic thrust generator. Although lift generation has been the focal point of most aerodynamic discussions on the origin of flight²², the importance of thrust has often been underexplored. Thrust, however, must have played a fundamental role in the origin of flight. As shown in our calculations, thrust is the only force that exerts work on *Archaeopteryx* along its entire take-off run (residual lift does not exert any work). Thus, we regard thrust, and not lift, as the primordial force ultimately responsible for sustained flight. Because the direction of thrust is perpendicular to that of gravity, not against it, objections to the cursorial theory on the basis of strenuous energetic demands² may not be relevant.

Even though our study centres on *Archaeopteryx*, our conclusions can be applied equally to non-avian maniraptoriforms flapping their wings in a downstroke–upstroke fashion. It is likely that these dinosaurs had the ability to passively supinate and flex their forelimbs^{9,18} as well as to flap them within ranges comparable to those of *Archaeopteryx*^{9,23}. Some of them have even been found to possess fully fledged wings²⁴. Thus, the structures and functions necessary for wing-generated thrust were already present in the flightless ancestors of birds. Long, vaned feathers, like those of the non-avian theropods *Caudipteryx* and *Protarchaeopteryx*²⁴, and the 'flight' stroke evolved in the context of terrestrial thrust. As previously implied^{25,26}, wing-generated thrust evolved before useful lift. Using this thrust and its ensuing residual lift, the flightless

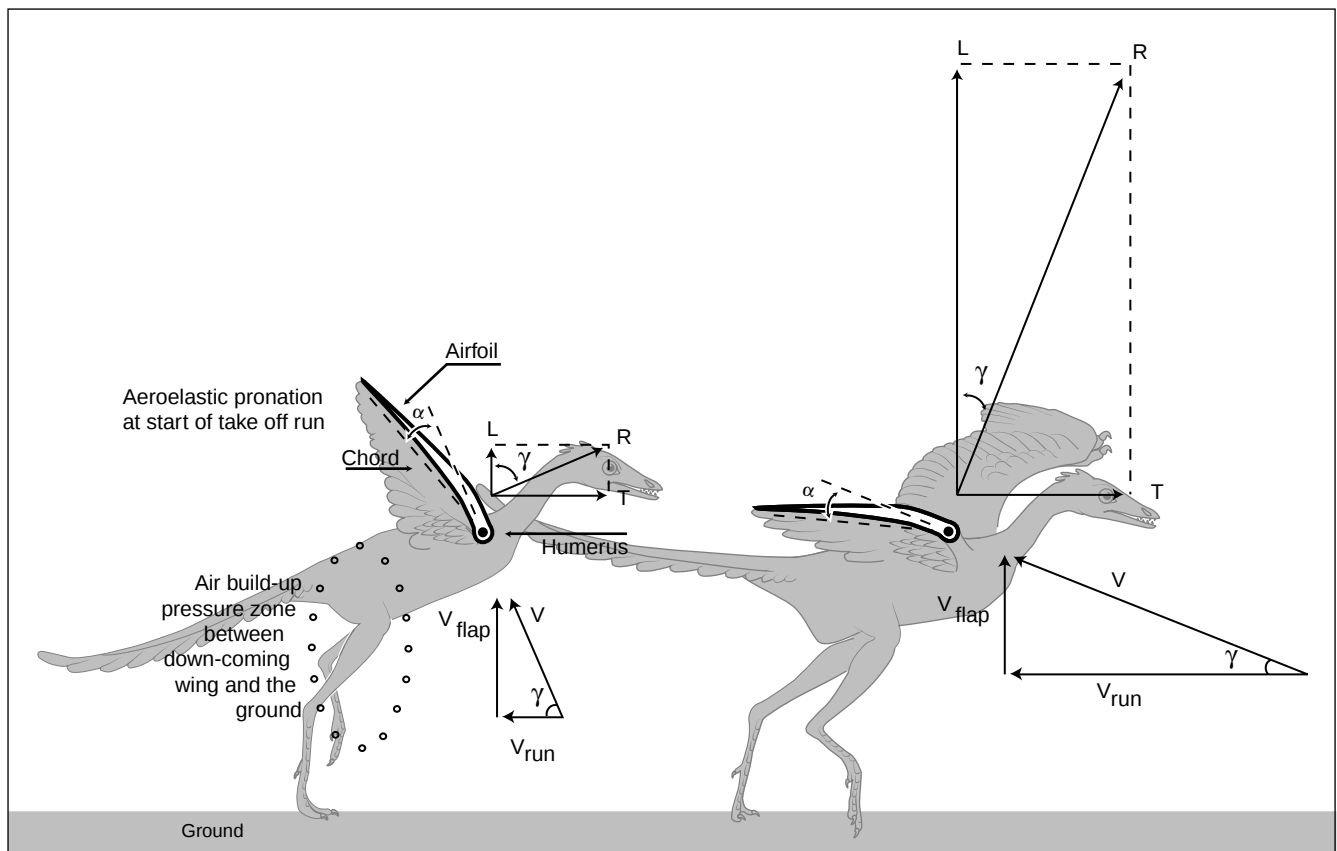


Figure 1 Vectorial representation (not to scale) of thrust generation and aeroelastic response at low running speeds (bold symbols represent vectors). While running at the speed V_{run} (left), *Archaeopteryx* flaps its wings downward with a velocity V_{flap} and encounters the upcoming flow V at a steep angle with the horizon (wing path angle γ). During the downstroke, the airfoil aligns itself to the upcoming airflow by rotating aeroelastically, trailing edge up, owing to the pressure build-up between the wing and the ground (elliptical dotted zone). This aeroelastic response is possible because the feathered wing of *Archaeopteryx*

is attached to the body only at the shoulder. This alignment is also expected as the result of wrist pronation along its semilunate carpal^{9,18}. Thus, the wing meets the incoming airflow V at an angle α and generates R , an aerodynamic force perpendicular to V . The horizontal and vertical projections of R are thrust T and lift L , respectively. As the running speed increases (right), the aerodynamic resultant R increases its magnitude and rotates counterclockwise. This rotating is due not to an induced drag increase but to an increase in the thrust and lift components. All forces shown are aerodynamic forces, not net forces on the bird.

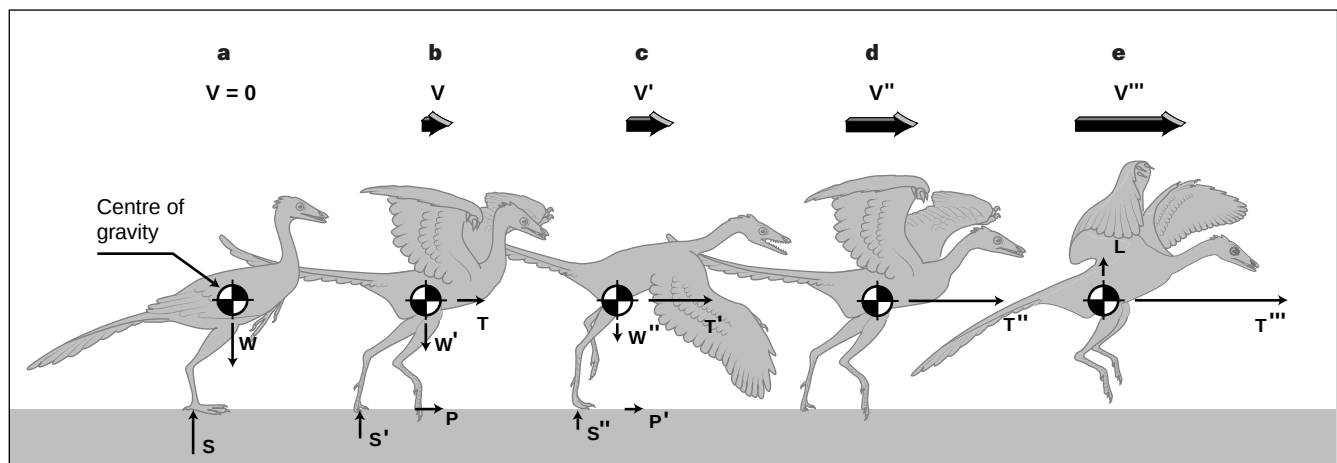


Figure 2 Net forces on *Archaeopteryx* throughout its modelled take-off run. The increasingly larger vectors (V to V''') depict the incremental velocity of *Archaeopteryx*. **a**, At standstill *Archaeopteryx* weighs W and the hindlimbs provide an equal and opposite force S to counteract it. **b**, At the beginning of the take-off run, *Archaeopteryx*'s hindlimbs produce a forward propulsion P , and when it starts to flap its wings these produce a thrust T . Owing to the generation of residual lift L , the hindlimbs need to supply a smaller upward force S' to counter the smaller net weight W' . This is called the 'vertical force migration'. **c**, During the take-off run, thrust T' gradually overtakes the propulsion P' (W' and S' decrease as a result of the vertical force migration). The migration of the forward force from

propulsion to thrust is called the 'horizontal force migration'. **d**, An instant before lift off, the horizontal force migration is completed and propulsion P' has disappeared. The only net force acting on the bird is thrust T'' . The vertical force migration is also completed, and the residual lift now equals weight. **e**, At lift-off, residual lift transitions into useful lift L . L , now a vertical net force, exerts work on the bird, lifting it up; this force translates into an ascensional velocity. Thrust T''' continues to exert work on the bird. 'Residual lift' refers to a lift force that does not act as a net force and that, as long as it is smaller than the bird's weight is unable to exert any vertical work on the bird. Since it is not a net force, residual lift is not depicted in **a-d**.

Table 1 Data for the curves of residual lift and net thrust in Fig. 3

	$V_{\text{run}} (\text{m s}^{-1})$	Net thrust (N)	Residual lift (N)	Weight (N)
Instant after standstill	0.15	0.32	0.013	1.96
	1.43	0.34	0.15	1.96
	2.71	0.37	0.34	1.96
Halfway through take-off run	3.99	0.45	0.61	1.96
	5.28	0.52	0.96	1.96
	6.56	0.60	1.41	1.96
At lift-off	7.84	0.66	1.96	1.96

Values for running velocity V_{run} , net thrust, residual lift and weight are given for an instant after standstill, half through the take-off run and at lift-off. Flapping frequency, 9.3 Hz. See Methods for equations.

ancestors of birds increased their cursorial velocity and ability to jump great heights.

Explanations of flight origins are conjectural and, as such, unlikely ever to be tested. The origin of bird flight from cursorial theropods is, however, not only the least conflicted hypothesis given the available phylogenetic and functional data but, as illustrated here, is also aerodynamically achievable. □

Methods

Our parameters for *Archaeopteryx* are those given by Rayner^{8,16} and Yalden²⁷; our mathematical model agrees with those used in take-off studies of large extant birds²⁰. In our model, an *Archaeopteryx* weighing 1.96 N (0.2 kg) runs and flaps its wings at 9.3 Hz. The wings accelerate uniformly downwards. The final vertical velocity of the wings (at the end of the downstroke) is calculated by $V_f = f_s \phi b y$, where f_s is the sum of wing strokes per second (18.6) (each cycle consisting of a downstroke and an upstroke), ϕ is the angle subtended by the wings during the downstroke (50°), b is the wingspan (0.58 m), and y is the span-wise wing station at which the vectorial calculations are performed (0.7 or 70% of semispan). The average flapping velocity of the wing V_{flap} is half the value of V_f . To calculate the relative air velocity across the airfoil V , the Pythagorean theorem is applied to add V_{run} and V_{flap} vectorially (Fig. 1). Throughout the calculations, V_{run} is considered to be an independent variable and ranges between 0.15 m s^{-1} and 7.84 m s^{-1} (take-off speed; headwinds would reduce this value) (Fig. 3; Table 1). The wing path angle γ is calculated by the arctangent of $V_{\text{flap}}/V_{\text{run}}$. The resultant aerodynamic force vector generated by the wing is calculated by $\mathbf{R} = (0.5)\rho V^2 \text{CL}S$ where ρ is the air density, V is the velocity, CL is the lift coefficient (2), S is the wing area (0.0479 m^2), and p is an average factor that considers lift being generated only during the downstroke (0.5). The relatively high value for lift coefficient (for the low Reynolds number of the wing) adopted here may be due to the combined effects of the proximity

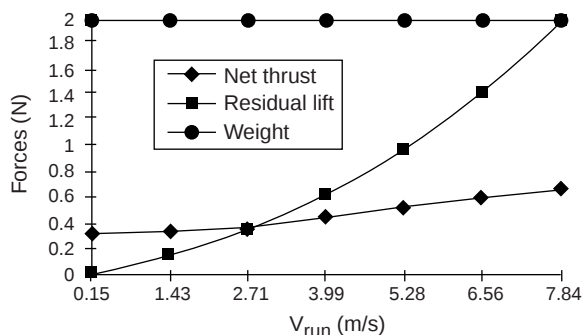


Figure 3 Progression of residual lift and net thrust throughout the modelled take-off run of *Archaeopteryx*. An instant after starting the take-off run, the thrust generated by flapping is 24 times higher than the residual lift (Table 1). Although for much of the take-off run the lift exerted on the bird is higher than thrust, thrust is the only force exerting work on the bird and increasing continuously the level of its kinetic energy. The cumulative power exerted by the thrust on the bird at a certain time during the take-off run can be calculated by the area under the thrust curve, from the beginning of the run until that time. At lift-off (the intersection of the residual lift and the weight curve), *Archaeopteryx* has reached the kinetic energy necessary for flight; residual lift cancels out weight. Residual lift becomes useful lift at lift-off.

of starting vortices on the upper side of the wing favouring pressure gradients, enhanced leading edge vortex lift, ground effect, and lift hysteresis²⁸. Given that vector $\mathbf{R}$ is always perpendicular to the incoming flow $\mathbf{V}$, its inclination $\mathbf{L}$ is also known (γ with respect to the horizontal). Thrust $\mathbf{T}$ and residual lift $\mathbf{L}$ are calculated by the horizontal and vertical projection of $\mathbf{R}$, respectively. The mathematical basis for the vertical force migration is that, during the take-off run, the sum of all vertical forces on the bird must be zero, because no vertical acceleration exists.

Received 15 May 1998; accepted 26 February 1999.

- Padian, K. & Chiappe, L. M. The early evolution of birds. *Biol. Rev.* **73**, 1–42 (1998).
- Norberg, U. M. *Vertebrate Flight* (Springer, Berlin, 1990).
- Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, 1996).
- Chatterjee, S. *The Rise of Birds* (John Hopkins Univ. Press, Baltimore, 1997).
- Hecht, M. K. et al. *The Beginnings of Birds* (Freunde Jura-Museum, Eichstätt, 1985).
- Padian, K. *The Origin of Birds and the Evolution of Flight* (Cal. Acad. Sci., San Francisco, 1986).
- Gauthier, J. Saurischian monophyly and the origin of birds. *Mem. Cal. Acad. Sci.* **8**, 1–55 (1986).
- Rayner, J. M. V. in *The Beginnings of Birds* (eds Hecht, M. K. et al.) 289–292 (Freunde Jura-Museum, Eichstätt, 1985).
- Gauthier, J. & Padian, K. in *The Beginnings of Birds* (eds Hecht, M. K. et al.) 185–197 (Freunde Jura-Museum, Eichstätt, 1985).
- Padian, K. in *Paleobiology* (eds Briggs, D. E. G. & Crowther, P. R.) 75–79 (Blackwell, Oxford, 1990).
- Rayner, J. M. V. in *Biomechanics in Evolution* (eds Rayner, J. M. V. & Wootton, R. J.) 183–212 (Cambridge Univ. Press, Cambridge, 1991).
- Ostrom, J. H. Bird flight: how did it begin? *Am. Sci.* **67**, 46–58 (1979).
- Peters, D. S. & Gorgner, E. A comparative study on the claws of *Archaeopteryx*. *L.A. Co. Mus., Sci. Ser.* **36**, 29–37 (1992).
- Chiappe, L. M. Climbing *Aechaeopteryx*? A Response to Yalden. *Archaeopteryx* **15**, 109–112 (1997).
- Poore, S. O., Ashcroft, A., Sánchez-Haiman, A. & Goslow, G. E. Jr The contractile properties of the M. supracoracoideus in the pigeon and starling: a case for long-axis rotation of the humerus. *J. Exp. Biol.* **200**, 2987–3002 (1997).
- Rayner, J. M. V. in *The Beginnings of Birds* (eds Hecht, M. K. et al.) 279–288 (Freunde Jura-Museum, Eichstätt, 1985).
- Jenkins, F. A. The evolution of the avian shoulder joint. *Am. J. Sci.* **293**, 253–267 (1993).
- Ostrom, J. H. How did flight might have come about. *Dinofest Inter. Proc.* 301–310 (1997).
- Speakman, J. R. Flight capabilities in *Archaeopteryx*. *Evolution* **47**, 336–340 (1993).
- Norberg, R. A. & Norberg, U. M. Take-off, landing, and flight speed during fishing flights of *Gavia stellata* (Pont.). *Ornis Scand.* **2**, 55–67 (1971).
- Ruben, J. Reptilian physiology and the flight capacity of *Archaeopteryx*. *Evolution* **45**, 1–17 (1991).
- Caple, G. et al. The physics of leaping animals and the evolution of preflight. *Am. Nat.* **121**, 455–476 (1983).
- Novas, F. E. & Puerta, P. F. New evidence concerning avian origins from the Late Cretaceous of Patagonia. *Nature* **387**, 390–392 (1997).
- Ji, Q., Currie, P. J., Norell, M. A. & Ji, S. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
- Nopcsa, F. Ideas on the origin of flight. *Proc. Zool. Soc. Lond.* **1907**, 223–236 (1907).
- Nopcsa, F. On the origin of flight in birds. *Proc. Zool. Soc. Lond.* **1923**, 463–477 (1923).
- Yalden, D. W. The flying ability of *Archaeopteryx*. *Ibis* **113**, 349–356 (1971).
- Burgers, P. A. *Computational Analysis of the Aerodynamic Interference Between Two Birds Flying in Formation and in Ground Effect Simultaneously*. Thesis, The Union Institute, Cincinnati, Ohio (1996).

Acknowledgements. We are grateful to H. Bertsch, A. Brush, A. Gaunt, P. MacCready, K. Padian, J. Ostrom, G. Spedding, and M. Warner for discussions, to P. Rollings for the illustrations and to S. Orell for editorial assistance. J. M. V. Rayner provided insightful comments on the earlier versions of the manuscript. Support provided by the Frank M. Chapman Memorial Fund of the American Museum of Natural History.

Correspondence and requests for materials should be addressed to P.B.

Induction and organization of Ca^{2+} waves by enteric neural reflexes

Randel J. Stevens, Nelson G. Publicover & Terence K. Smith

Biomedical Engineering Program and Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, Nevada 89557, USA

The motility of the gastrointestinal tract consists of local, non-propulsive mixing (pendular or segmental) and propulsive (peristaltic) movements^{1–5}. It is generally considered that mixing movements are produced by intrinsic pacemakers which generate rhythmic contractions^{4–6}, and peristalsis by intrinsic excitatory and inhibitory neural reflex pathways^{1–5,7,8}, but the relationship between mixing and peristalsis is poorly understood^{4–6}. Peristalsis is compromised in mice lacking interstitial cells of Cajal⁹, suggesting that these pacemaker cells^{10–14} may also be involved in neural reflexes. Here we show that mixing movements within long-

itudinal muscle result from spontaneously generated waves of elevated internal calcium concentration which originate from discrete locations (pacing sites), spread with anisotropic conduction velocities in all directions, and terminate by colliding with each other or with adjacent neurally suppressed regions. Excitatory neural reflexes control the spread of excitability by inducing new pacing sites and enhancing the overall frequency of pacing, whereas inhibitory reflexes suppress the ability of calcium waves to propagate. We provide evidence that the enteric nervous system organizes mixing movements to generate peristalsis, linking the neural regulation of pacemakers to both types of gut motility.

By measuring the fluorescent intensity of the Ca^{2+} indicator Fluo-3, we visualized the origin and spread of excitation in longitudinal muscles of isolated, intact colonic tissues (Fig. 1a). In all preparations, spontaneous waves of increased internal calcium concentration (frequency = $1.3 \pm 0.5 \text{ s}^{-1}$; $n = 50$ tissues) propagated with anisotropic conduction velocities through the longitudinal muscle layer (Fig. 1a). Propagation velocities of Ca^{2+} waves were approxi-

mately 10 times faster when parallel ($138 \pm 33 \text{ mm s}^{-1}$; vertical direction in the field of view in Fig. 1a; $n = 14$ tissues) than when orthogonal ($14.1 \pm 3.1 \text{ mm s}^{-1}$; $n = 21$ tissues) to the long axis of longitudinal muscle fibres (Fig. 1a). Spontaneous Ca^{2+} waves usually emerged from one (Fig. 1a) or at most two (Fig. 1d) discrete sites and consistently spread beyond the field of view ($1.2 \times 1.0 \text{ mm}$) parallel to the longitudinal muscle fibres and for varying distances across the circumferential axis. They emerged within the field of view (Fig. 1a, d), or entered without preference from the oral, anal or circumferential directions (Fig. 1b). These waves occasionally stopped abruptly (Fig. 1c, wave v) at locations that had previously supported propagation (Fig. 1c, waves i, ii and vi).

Spontaneous Ca^{2+} waves appear to underlie mixing movements, because propagation through regions of muscle in loosely pinned preparations generated local contractions. Ca^{2+} waves were completely blocked by the L-type Ca^{2+} -channel antagonist nifedipine ($1 \mu\text{M}$), like the spontaneous rhythmic contractions and their

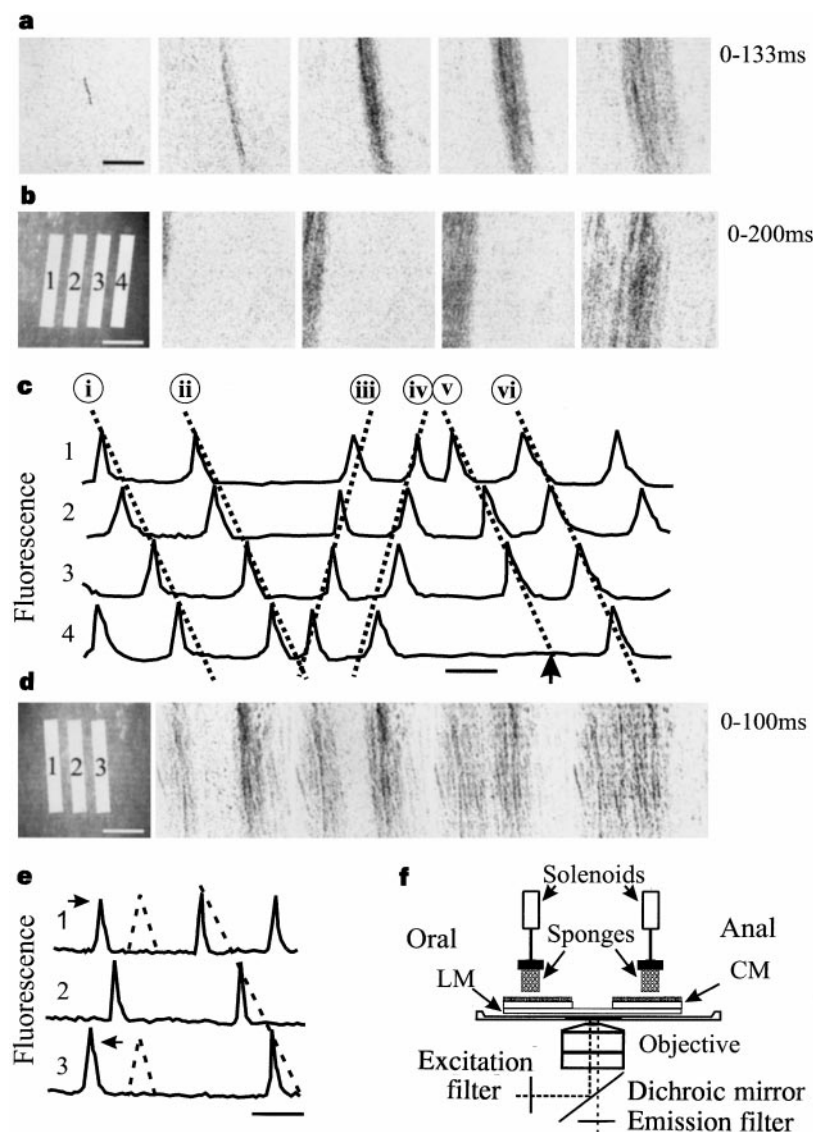


Figure 1 Visualization of spontaneous Ca^{2+} -waves. **a**, Sequential difference images show a Ca^{2+} wave emerging from a pacing site. Summed fluorescent intensities from four regions (20×320 pixels) indicated in the left frame of **b** (background image; calibration bar, $300 \mu\text{m}$) are shown in **c** (time bar, 0.5 s). Ca^{2+} waves spread circumferentially from distinct pacing sites both left-to-right (**c**, waves i, ii, v and vi) and right-to-left (**c**, waves iii and iv). One wave stopped abruptly (**c**, wave v, trace 4) as a new wave emerged spontaneously (**c**, wave vi, trace 1). Summed fluorescent intensities (regions indicated in first frame of **d**; 20×300 pixels; calibration bar, $300 \mu\text{m}$) shown in **e** (time bar, 1 s) demonstrate a collision of two events. Arrows illustrate the direction of colliding Ca^{2+} waves. Dashed wave forms indicate the arrival times of Ca^{2+} waves that would be expected if they had continued with the same conduction velocities after the collision. **f**, Schematic of the preparation, imaging apparatus and mucosal stimulators. CM, circular muscle; LM, longitudinal muscle.

trace 1). Summed fluorescent intensities (regions indicated in first frame of **d**; 20×300 pixels; calibration bar, $300 \mu\text{m}$) shown in **e** (time bar, 1 s) demonstrate a collision of two events. Arrows illustrate the direction of colliding Ca^{2+} waves. Dashed wave forms indicate the arrival times of Ca^{2+} waves that would be expected if they had continued with the same conduction velocities after the collision. **f**, Schematic of the preparation, imaging apparatus and mucosal stimulators. CM, circular muscle; LM, longitudinal muscle.

associated Ca^{2+} transients in the longitudinal muscle of the guinea-pig ileum¹⁵. Ca^{2+} waves are probably generated by action potentials superimposed upon electrical slow waves initiated by interstitial cells of Cajal^{16,17}. Conduction velocities appear to be regulated by the passive electrical properties of the longitudinal muscle, which exhibit a similar anisotropy¹⁵. The frequency of Ca^{2+} waves is similar to the frequency of spontaneous action potentials in colonic longitudinal muscle^{18–20}. Video images show that irregularities in the periodicity (arrhythmias) of Ca^{2+} transients at any one site result from shifts in the locations of pacing sites and differences in conduction times to the recording site (Fig. 1c).

It is important to understand how Ca^{2+} waves originating from different pacing sites interact to modify the contractile patterns of longitudinal muscle. The collision of two circumferentially propagating Ca^{2+} waves (Fig. 1d and e, traces 1 and 3) always resulted in a unitary event of similar amplitude to either wave (Fig. 1e, trace 2), and failed to propagate in either direction beyond the site of collision (Fig. 1e, dashed waves). Collisions between Ca^{2+} waves in the circumferential direction did not appear to affect propagation in the longitudinal axis. The collisions resemble those between

action potentials evoked to propagate in opposite directions along nerves, in that Ca^{2+} waves are followed by a trail of refractoriness which prevents them from emerging from a collision site. Following excitation of a region by a Ca^{2+} wave, muscles remained unable to support propagation of subsequent events for 133 ± 33 ms (minimum inter-event interval; 30 events in 8 tissues). Refractoriness is likely to be a consequence of inactivation of the conductances associated with action potentials that underlie the Ca^{2+} waves. The maximum frequency of activity sustainable in a given region is governed by this refractory interval as well as by the availability of events from pacing sites in the vicinity.

Peristalsis is organized by the activation of ascending and descending neural reflex pathways that interact to produce a combined wave of contraction preceded by relaxation^{1,2,7,8,21}, which occurs synchronously in both the longitudinal and circular muscle layers^{22,23}. Previous studies have shown that there is an increase in the strength and rate of pulsatile contractions and an increase in the frequency of action potential firing oral to a stimulus²⁴. The effects of activating oral reflex pathways were characterized by stimulating the mucosa anal to the recording location (ascending

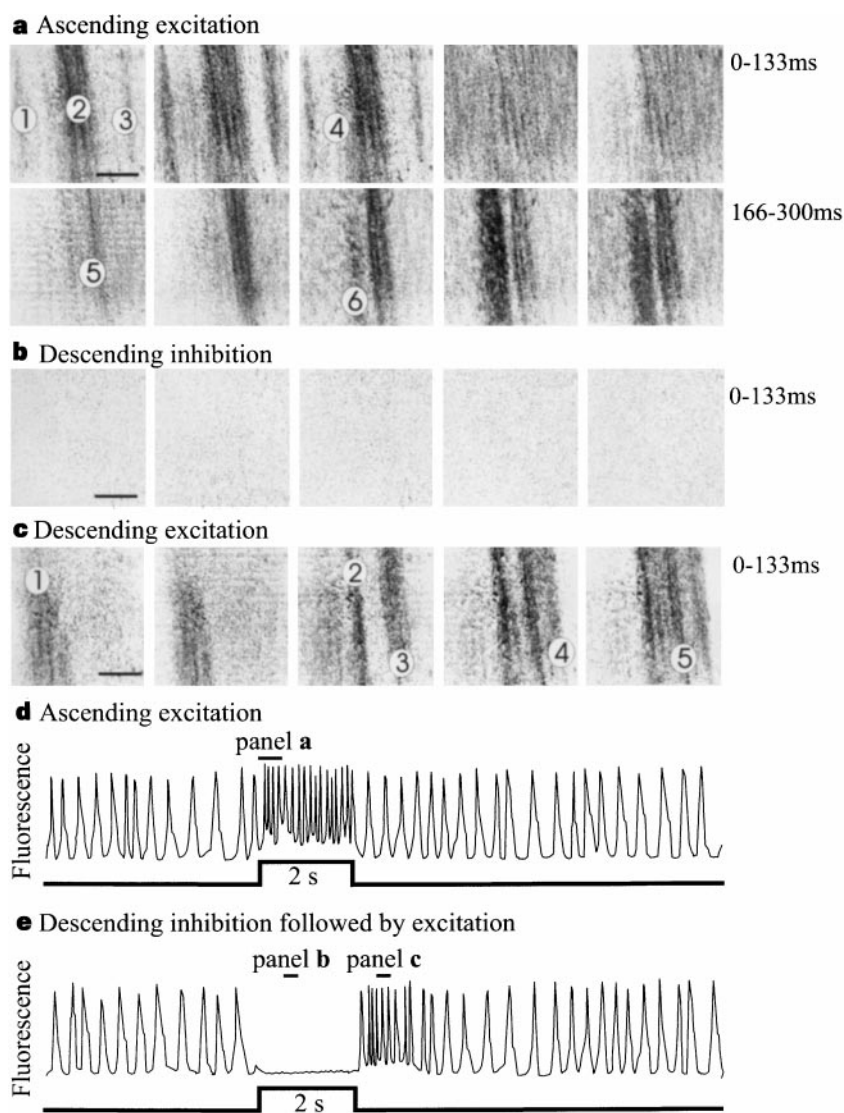


Figure 2 Neural reflex responses. Sequential difference images show induction (a, c) and suppression (b) of pacing sites in longitudinal muscle immediately following the application of a mucosal stimulus applied anal or oral (indicated by time bars in d and e, respectively) to the recording site (calibration bars, 300 μm).

Summed pixel intensities over time from a central region (20 $\times$ 320 pixels) illustrate the effects of mucosal stimulation (lower trace) applied anal (d) or oral (e) to the recording site. Frames in a, b and c were sampled at times indicated by bars above traces in d and e.

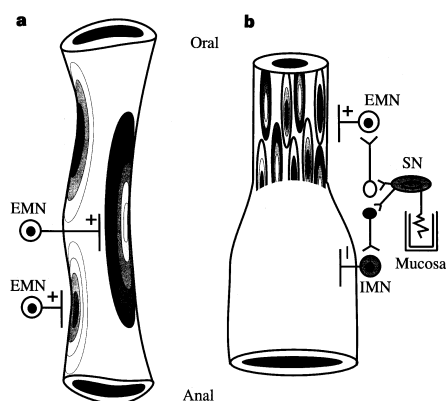


Figure 3 Neurally organized mixing patterns lead to peristalsis. **a**, Mixing is regulated by relatively few pacing sites which propagate over large regions, creating shearing forces and tone. **b**, Peristalsis consists of activation of polarized nerve pathways, which synchronously induce additional pacing sites above the region of mucosal stimulation to promote intestinal shortening and suppressed propagation of excitability below the stimulus to allow distension. EMN, excitatory motor neuron; IMN, inhibitory motor neuron; SN, sensory neuron located in the mucosa.

excitation; Fig. 1f)^{23,25}. The activation of ascending neural reflex pathways induced multiple pacing sites (from 1.4 ± 0.5 to 12.1 ± 2.3 sites per s within a 1.2 mm^2 field of view, $P < 0.001$, 12 tissues; Fig. 2a) and increased the overall frequency of Ca^{2+} transients observed at any single location from 1.3 ± 0.5 to 8.3 ± 2.2 waves per s (30 events in 6 tissues; Fig. 2d; ascending excitation).

During excitation, circumferential propagation of Ca^{2+} waves was limited by collisions between waves arising from adjacent pacing sites (Fig. 2a). Pacing-site density during stimulation corresponds to ~ 200 sites induced within a 1-mm band of muscle around the circumference of the intestine (~ 12 sites per s per mm^2 over a circumference of 18 mm). Collisions prevent any one pacing site from controlling large regions of muscle, enabling the neurally coordinated excitation of many small parallel bands of longitudinal muscle arranged side-by-side about the circumference of the colon (Fig. 3b; excited region). The minimum region activated by each pacing site is approximately the minimum width ($\sim 100 \mu\text{m}$) of a muscle bundle necessary to sustain action potential propagation^{26,27}. Ascending excitation induces rapid mixing activity synchronously at multiple sites around the wall of the gut to produce a uniform shortening of the intestine.

In contrast, activation of descending neural reflexes (oral mucosal stimulation; Fig. 1f) abolished spontaneous initiation and propagation of Ca^{2+} waves within the longitudinal muscle for the duration of the stimulus (Fig. 2b and e; descending inhibition). These descending inhibitory responses underlie the abolition of spontaneous contractions and relaxation reported by others^{5,7,22–24}. Following the termination of the oral stimulus, there was a period of after-excitation which induced multiple pacing sites (11.5 ± 0.5 sites per s within a field of view, 30 events in 6 tissues; Fig. 2c and e; descending excitation) and an increased overall frequency (from 1.3 ± 0.5 to 8.0 ± 0.7 waves per s, 6 tissues). Descending after-excitation re-excites the formerly inhibited region, providing further force to promote propulsion^{5,7,22,23}. All reflex responses were abolished by the sodium-channel blocker tetrodotoxin ($1 \mu\text{M}$; 12 tissues), suggesting that they are neural in origin. In addition, the muscarinic antagonist atropine ($1 \mu\text{M}$) completely abolished all excitatory reflex responses (7 tissues). Atropine also abolished spontaneous Ca^{2+} waves in four tissues; however, in the three other tissues, spontaneous Ca^{2+} waves returned at a suppressed frequency (0.2 ± 0.1 waves per s) after at least 10 min of exposure (total observation period 30 min) to the drug. These data

suggest that pacing sites can be induced by the release of acetylcholine from excitatory motor neurons.

In summary, our data show that, during mixing, Ca^{2+} waves emerge from discrete sites and spread anisotropically within longitudinal muscle. Factors that affect the size of a region controlled by an individual pacing site include: (1) collisions between Ca^{2+} waves originating from adjacent pacing sites; (2) termination of events when they encounter refractory regions; (3) events encountering neurally suppressed regions; and (4) morphological barriers. Mixing appears to be promoted by asymmetric shearing forces created by Ca^{2+} waves spreading around the intestine over relatively large distances (Fig. 3a). Neural organization of mixing (pendicular or segmental) movements by polarized reflex pathways leads to the oral excitatory and descending inhibitory components of peristalsis (Fig. 3b). Oral excitation is produced by the neural induction and increased frequency of pacing sites, which ensure that synchronous bands of longitudinal muscle around the intestine contract to provide uniform intestinal shortening. Descending relaxation is produced by the prevention of the emergence of local Ca^{2+} waves and suppression of propagation into the neurally inhibited zone from adjacent regions of excited longitudinal muscle (Fig. 3b). Enteric neural reflex pathways create and control the movement of an interface where adjoining excited and inhibited regions of longitudinal muscle co-exist to generate peristalsis. □

Methods

Segments of distal colon (4 cm) were removed from male guinea-pigs (200–300 g), opened and pinned flat with the mucosa uppermost^{23,28}. Longitudinal muscles (mucosa and circular muscle removed, myenteric plexus intact over sampling region; Fig. 1f) were loaded with the Ca^{2+} indicator Fluo-3 acetoxymethyl ester ($5 \times 10^{-6} \text{ M}$), 0.01% dimethyl sulphoxide and 0.025% cremophor EL for 35–45 min in Krebs ringer bicarbonate solution at 37°C . Muscles were illuminated at $470 \pm 20 \text{ nm}$ and fluorescent emissions $>510 \text{ nm}$ were recorded as a relative measure of intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) using an intensified camera (30 frames per s; 640×480 pixels). Series of enhanced difference images were computed by storing an initial frame (with no apparent activity) as a 'background image' (first frame of Fig. 1b and d). Each subsequent frame was subtracted from this image, resulting in dark areas that indicate regions of increased fluorescent intensity. To monitor Ca^{2+} transients over prolonged periods, regions of interest were selected and pixel intensities were summed, normalized for area and displayed as a function of time. Conduction velocities were measured by determining the distance travelled by a propagating wave front over each digitized image (33-ms interval). Ascending and descending neural reflexes were evoked by depressing the mucosa anal or oral to the recording site, respectively, with a sponge mounted on a computer-controlled solenoid (Fig. 1f). Statistics are reported as mean $\pm$ standard deviation.

Received 27 January; accepted 18 February 1999.

- Bayliss, W. & Starling, E. H. Movements and innervation of the small intestine. *J. Physiol. (Lond.)* **24**, 99–143 (1899).
- Bayliss, W. & Starling, E. H. The movements and innervation of the large intestine. *J. Physiol. (Lond.)* **26**, 107–118 (1900).
- Kosterlitz, H. W. & Robinson, J. A. Reflex contraction of the longitudinal muscle coat of the isolated guinea-pig ileum. *J. Physiol. (Lond.)* **146**, 369–379 (1959).
- Smith, T. & Sanders, K. in *Textbook of Gastroenterology* (eds Yamada, T., Alpers, D., Owyang, C. & Powel, D.) 234–260 (Lippincott, Silverstein, Philadelphia, 1995).
- Holman, M. in *Smooth Muscle and Assessment of Current Knowledge* (ed. Bulbring, E.) 311–338 (University of Texas Press, Austin, 1981).
- Christensen, J. in *Physiology of the Gastrointestinal Tract* Vol. 1 (ed. Johnson, L. R.) 991–976 (Raven, New York, 1994).
- Hirst, G. D. S., Holman, M. E. & McKirdy, H. C. Two descending nerve pathways activated by distension of guinea-pig small intestine. *J. Physiol. (Lond.)* **244**, 113–127 (1975).
- Mackenna, B. R. & McKirdy, H. C. Peristalsis in the rabbit distal colon. *J. Physiol. (Lond.)* **220**, 33–54 (1972).
- Huizinga, J. D., Ambros, K. & Der-Silaphet, T. Co-operation between neural and myogenic mechanisms in the control of distension-induced peristalsis in the mouse small intestine. *J. Physiol. (Lond.)* **506**, 843–856 (1998).
- Thomsen, L. et al. Interstitial cells of Cajal generate a rhythmic pacemaker current. *Nature Med.* **4**, 848–851 (1998).
- Ward, S. M., Harney, S. C., Bayguinov, J. R., McLaren, G. J. & Sanders, K. M. Development of electrical rhythmicity in the murine gastrointestinal tract is specifically encoded in the tunica muscularis. *J. Physiol. (Lond.)* **505**, 241–258 (1997).
- Huizinga, J. D. et al. W/kit gene required for interstitial cells of Cajal and for intestinal pacemaker activity. *Nature* **373**, 347–349 (1995).

13. Koh, S., Sanders, K. & Ward, S. Spontaneous electrical rhythmicity in cultured interstitial cells of Cajal from the murine small intestine. *J. Physiol. (Lond.)* **513**, 203–213 (1998).
14. Publicover, N. G. in *Pacemaker Activity and Intercellular Communication* (ed. Huizinga, J. D.) 175–192 (CRC, Boca Raton, 1995).
15. Cousins, H. M., Edwards, F. R., Hirst, G. D. & Wendt, I. R. Cholinergic neuromuscular transmission in the longitudinal muscle of the guinea-pig ileum. *J. Physiol. (Lond.)* **471**, 61–86 (1993).
16. Horowitz, B., Ward, S. & Sanders, K. Cellular and molecular basis for electrical rhythmicity in gastrointestinal muscles. *Annu. Rev. Physiol.* **61**, 19–43 (1999).
17. Ozaki, H., Stevens, R. J., Blondfield, D. P., Publicover, N. G. & Sanders, K. M. Simultaneous measurements of membrane potential, cytosolic Ca^{2+} , and tension in intact smooth muscles. *Am. J. Physiol.* **260**, C917–925 (1991).
18. Holman, M. E. Membrane potentials recorded with high resistance microelectrodes and the effects of changes in ionic environment on the electrical and mechanical activity of the smooth muscle of the taenia coli of the guinea pig. *J. Physiol. (Lond.)* **141**, 464–488 (1958).
19. Gillespie, J. S. Spontaneous mechanical and electrical activity of stretched and unstretched intestinal smooth muscle cells and their response to sympathetic nerve stimulation. *J. Physiol. (Lond.)* **162**, 54–75 (1962).
20. Smith, T. K., Reed, J. B. & Sanders, K. M. Interaction of two electrical pacemakers in muscularis of canine proximal colon. *Am. J. Physiol.* **252**, C290–299 (1987).
21. Costa, M. & Furness, J. B. The peristaltic reflex: an analysis of the nerve pathways and their pharmacology. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **294**, 47–60 (1976).
22. Smith, T. K. & Robertson, W. J. Synchronous movements of the longitudinal and circular muscle during peristalsis in the isolated guinea-pig distal colon. *J. Physiol. (Lond.)* **506**, 563–577 (1998).
23. Smith, T. K. & McCarron, S. L. Nitric oxide modulates cholinergic reflex pathways to the longitudinal and circular muscle in the isolated guinea-pig distal colon. *J. Physiol. (Lond.)* **512**, 893–906 (1998).
24. Costa, M. & Furness, J. B. in *Mediators and Drugs in Gastrointestinal Motility I. Morphological Basis of Neurophysiological Control* (ed. Bertacchini, A.) 279–382 (Springer, Berlin, 1982).
25. Smith, T. K. & Furness, J. B. Reflex changes in circular muscle activity elicited by stroking the mucosa: an electrophysiological analysis in the isolated guinea-pig ileum. *J. Auton. Nerve Syst.* **25**, 205–218 (1988).
26. Tomita, T. Electrical responses of smooth muscle to external stimulation in hypertonic solution. *J. Physiol. (Lond.)* **183**, 450–468 (1966).
27. Burnstock, G. & Prosser, C. Conduction in smooth muscle, comparative electrical properties. *Am. J. Physiol.* **199**, 553 (1960).
28. Smith, T. K., Bornstein, J. C. & Furness, J. B. Convergence of reflex pathways excited by distension and mechanical stimulation of the mucosa onto the same myenteric neurons of the guinea pig small intestine. *J. Neurosci.* **12**, 1502–1510 (1992).

Acknowledgements. This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases.

Correspondence and requests for materials should be addressed to T.K.S. (e-mail: tks@physio.unr.edu).

Dendritic spine changes associated with hippocampal long-term synaptic plasticity

Florian Engert & Tobias Bonhoeffer

Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, 82152 München-Martinsried, Germany

Long-term enhancement of synaptic efficacy in the hippocampus is an important model for studying the cellular mechanisms of neuronal plasticity, circuit reorganization, and even learning and memory¹. Although these long-lasting functional changes are easy to induce, it has been very difficult to demonstrate that they are accompanied or even caused by morphological changes on the subcellular level. Here we combined a local superfusion technique^{2,3} with two-photon imaging⁴, which allowed us to scrutinize specific regions of the postsynaptic dendrite where we knew that the synaptic changes had to occur. We show that after induction of long-lasting (but not short-lasting) functional enhancement of synapses in area CA1, new spines appear on the postsynaptic dendrite, whereas in control regions on the same dendrite or in slices where long-term potentiation was blocked, no significant spine growth occurred.

We used a local superfusion technique^{2,3} to investigate whether long-term potentiation (LTP) of synaptic efficacy is accompanied by subcellular morphological changes. This allowed us to restrict synaptic activity and thus the site of LTP induction to a small region on the dendrite of the postsynaptic neuron. After a pyramidal neuron in area CA1 of a hippocampal slice culture was impaled with an intracellular recording electrode (resting potential of less than -65 mV) and filled with the fluorescent dye calcein, we evoked excitatory postsynaptic potentials (EPSPs) by electrical stimulation

of the Schaffer collaterals (Fig. 1a). We then blocked transmitter release everywhere but in a small area of roughly $30\text{ }\mu\text{m}$ diameter, in which the blocking solution ($10\text{ }\mu\text{M Cd}^{2+}$, 0.8 mM Ca^{2+}) was displaced by 'normal' superfusion solution (Fig. 1b; see Methods). After we had identified an area with a suitable group of synapses, baseline synaptic transmission was recorded. We then induced LTP by pairing postsynaptic depolarizations with single presynaptic stimuli. Before, during and after the pairing, we used a two-photon laser-scanning microscope to acquire high-resolution three-dimensional image stacks of the postsynaptic neuron (Fig. 1c) with a frequency of usually six stacks per hour. Figure 2 shows a representative example of a dendritic branch located under the superfusion spot before, during and after the induction of LTP. The electrophysiological record (Fig. 2a) demonstrates the enhance-

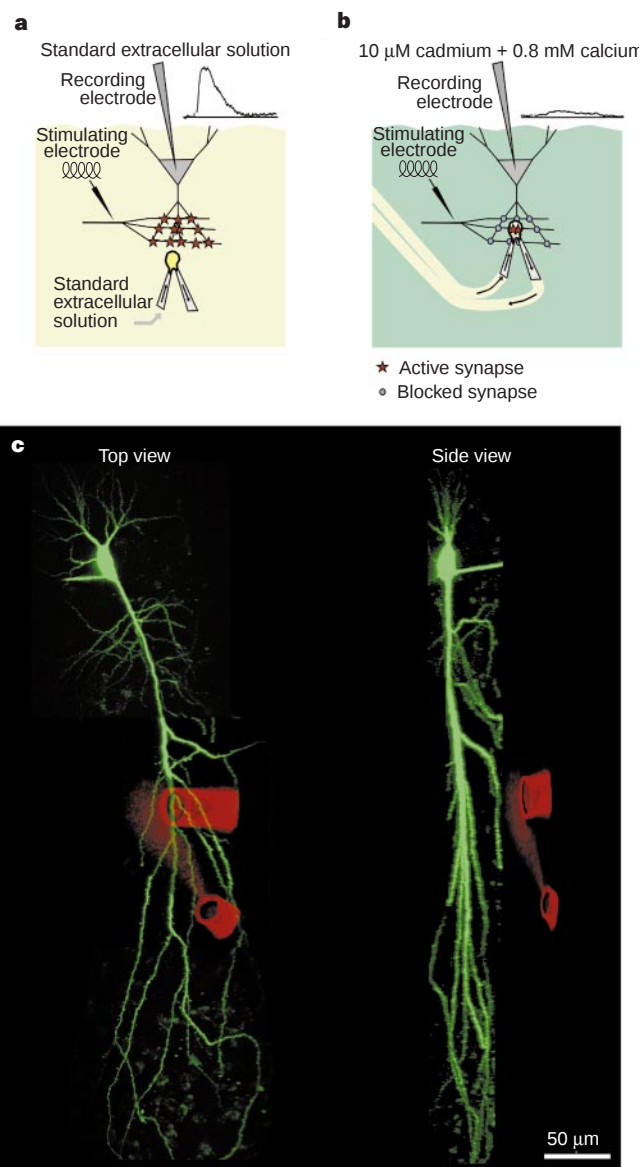


Figure 1 Experimental setup. **a, b**, Drawing of the recording situation. Actual EPSPs are shown as insets in the top right-hand corners. Active synapses are depicted as red stars, synapses blocked by $10\text{ }\mu\text{M Cd}^{2+}$ and low Ca^{2+} as blue dots. **c**, Two-photon laser-scanning image of a CA1 pyramidal neuron with the superfusion spot and the two pipettes (imaged on a different channel of the microscope with phase-contrast illumination) superimposed in false colours. Left, a maximum intensity projection through the complete picture stack; right, the same stack rotated by 90° . Only the dendritic branch that extends up to the surface of the slice is reached by the superfusion spot.

ment of synaptic strength. The columns of pictures for the different time points show the dendrite of interest from three different viewing angles (Fig. 2b). Clearly, two spines appear after synaptic enhancement in the positions marked with arrowheads. No spines can be seen at these positions in the rotated images of the first column, indicating that the appearance of spines cannot be attributed to a rotation artefact.

The raw data ('maximum intensity projection') of this (Fig. 3a) and the six remaining experiments in which LTP was successfully induced are shown in Fig. 3. In the first column, red dots on the

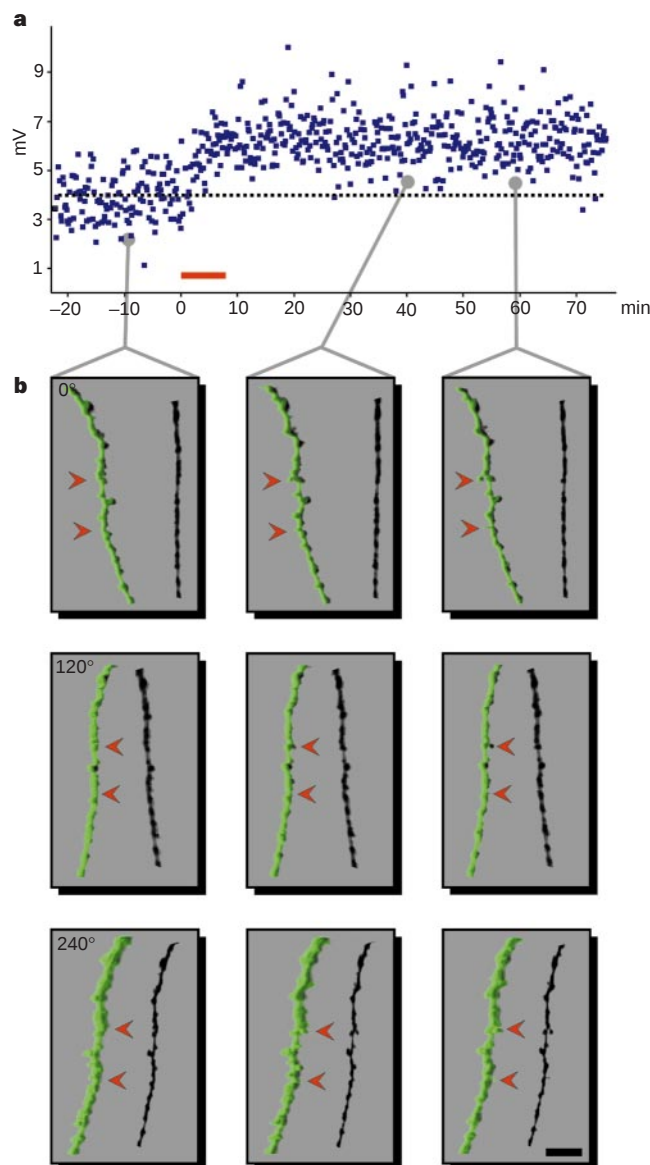


Figure 2 New spines emerge after the induction of LTP. **a**, Raw electrophysiological data. EPSP amplitudes of a pharmacologically isolated group of synapses are shown over time. After the pairing protocol, which was applied at the time indicated by the red bar, a stable and long-lasting potentiation of synaptic efficacy can be observed. **b**, Three columns of pictures taken at the time points indicated by the grey lines show a stretch of dendrite within the superfused region 'volume-rendered' from three different viewing angles (0°, 120° and 240°). A set of spines appears roughly 30 min after LTP induction at the sites marked by red arrowheads. The volume-rendering from different perspectives shows that the spines had really appeared *de novo* and were not just hidden behind the dendrite in the initial pictures. Scale bar, 5 μ m. For animated time-lapse data, see Supplementary Information.

neuron show the positions at which new spines grew, and blue dots denote where spines were found to disappear (see Methods, Blind evaluation). New spines usually appeared only under the superfusion spot (green overlay). Disappearing spines (blue dots) are seen more rarely, but at all locations. The panels to the right show magnifications of the regions of interest at -10, +30 and +60 min, and at an additional variable time point relative to LTP induction, as well as the electrophysiological data.

Three different conditions were used as controls: (1) regions far outside the superfusion spot ('off spot') were scrutinized for morphological changes. Apart from one case (Table 1), we saw no new spines in these areas (an example is shown in Fig. 4a), while disappearing spines were still observed; (2) experiments in which the induction of LTP was prevented by the NMDA (N-methyl-D-aspartate)-receptor antagonist D,L-2-amino-5-phosphonovalerate (AP5, 50 μ M); and (3) experiments in which the pairing procedure did not result in a stable potentiation were analysed. In the last two cases, we never observed new spines after the pairing (Table 1 and Fig. 4b, c) but spines still disappeared in 'random' locations (Table 1). Perhaps the most interesting control is one case in which the pairing resulted only in a short potentiation of synaptic efficacy (<30 min; data not shown). Also here we did not observe any new spines.

For an unbiased analysis, an independent, 'blind' observer (see Methods) inspected the image stacks 'section by section' for microscopic morphological changes in the imaged neuron. In every experiment in which LTP was successfully induced ($n = 7$), the blind observer detected between two and nine new spines in the region underneath the superfusion spot (Table 1, LTP). We quantified the emergence or disappearance of spines by calculating the spine density on a length of dendrite (spines per 100 μ m) before and after the induction procedure (Table 1). Although most spines remained stable, dendrites located under the superfusion spot showed a highly significant correlation between the appearance of new spines and an increase in synaptic efficacy (Fig. 5, red bars; Table 1; $P < 0.001$, two-tailed Students t -test). The disappearance of spines, on the other hand, was slightly anticorrelated with the successful induction of LTP (Fig. 5, blue bars; $P < 0.05$, two-tailed Students t -test).

Activity-related changes in dendritic structure and spine density have been examined in several preparations, but most of these studies have focused on long-term changes over several days to months⁵⁻⁸. Whether rapid changes in synaptic efficacy, such as those in LTP or LTD, are also accompanied by structural modifications is still a controversial question⁹⁻¹². The general difficulty linked to this issue is twofold: first, the exact spatial location of the synapses in which LTP is induced is usually not known, so thousands of spines have to be investigated^{6,10}; and second, statistical comparison of treated with untreated preparations requires a very large sample number to tease out the relatively small morphological changes⁶. A

Table 1 Statistics of the complete data set

Type of experiment	LTP	AP5	no LTP	'off spot'
Number of experiments	7	3	2	3
Length of dendrite examined (μ m)	537	468	329	2,093
Spines counted	216	211	182	1,052
Spines per dendrite (μ m ⁻¹)	0.40	0.45	0.55	0.50
New spines	32	0	0	4
100 $\times$ new spines per dendrite (μ m ⁻¹)	6.0	0	0	0.2
Lost spines	1	11	0	26
100 $\times$ lost spines per dendrite (μ m ⁻¹)	0.2	2.3	0	1.2

The first column shows data from experiments in which LTP was induced successfully. The last three columns show data from three different controls: block of LTP by the NMDA-antagonist AP5; unsuccessful pairing (no LTP); and data from outside the superfusion spot ('off spot'). For calculation of the lengths of dendrites the third dimension of the picture stack was taken into consideration; that is only the superficial 30 μ m were ascribed to the LTP group whereas the rest of each respective experiment was pooled with the 'off spot' control (see Methods). Note that there are only four 'inappropriate' new spines in the off-spot column which were detected by the blind observer in a small cluster in only one of the experiments.

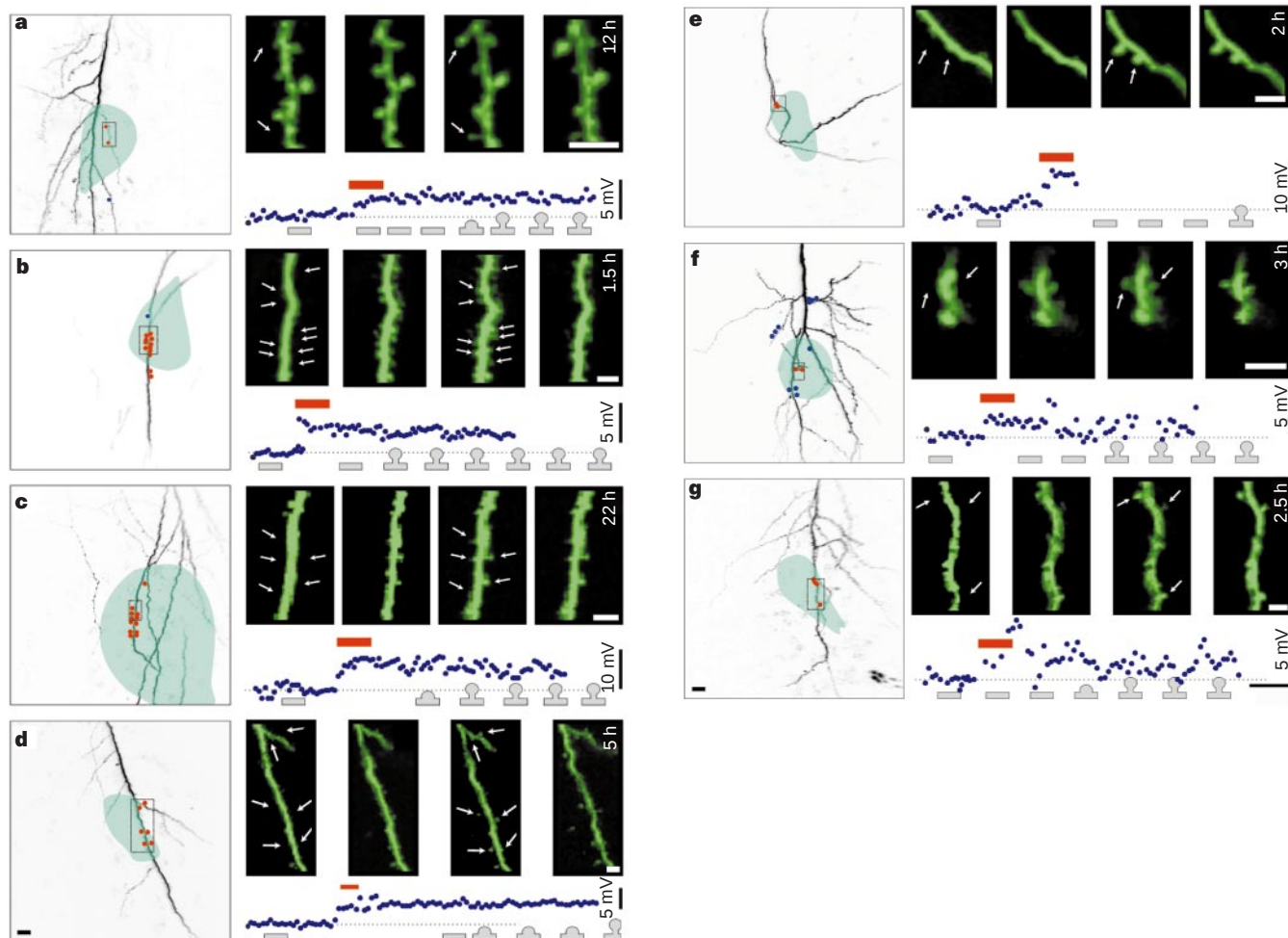


Figure 3 Experiments that produced a change in synaptic efficacy. **a–g**, Complete data set of experiments. The first experiment (**a**) is the same as that shown in Fig. 2. Left, a $170\ \mu\text{m} \times 170\ \mu\text{m}$ maximum intensity projection of part of the dendrite is shown with the superfusion spot superimposed in false colour. Red dots denote the points where new spines were found to emerge; blue dots show points where spines disappeared. Note that the disappearing spines in **f** are located deep in the slice culture ($>50\ \mu\text{m}$) and therefore stem from an 'off-spot' region. The first three right-hand columns show the magnified area of interest (dashed rectangle) at three time points: -10 , $+30$ and $+60$ min; the last shows varying later time points as indicated in the panels (up to 22 h). White

further complication is that if positive and negative changes happen simultaneously, as we show they do, and if they occur with about the same frequency, both kinds of change will go undetected. Therefore, the combination of the local superfusion technique (reducing the potential area for morphological changes to a region of approximately $30\ \mu\text{m}$ diameter) and two-photon laser scanning (reducing the phototoxic damage⁴ and thus greatly facilitating observations in living tissue) was crucial for our ability to detect morphological changes associated with LTP.

It has been proposed that the strength of synapses is controlled by changes in spine shape^{9–11,13–16}, such as shortening and/or widening of the neck which would reduce its electrical resistance and thus increase synaptic efficacy¹⁴. The results presented here do not directly address this issue; rather, they demonstrate that the growth of entirely new spines correlates with the induction of LTP. We cannot rule out, however, that additional smaller changes in spine morphology below our spatial resolution may also co-vary with synaptic strength.

We also have no direct proof that the emerging spines contain active synapses. However, most of the spines on pyramidal cells

arrows indicate positions at which new spines were detected by blind analysis. They are only displayed in the first and third panels, which show the data used for the blind analysis. Below the four panels, electrophysiological data from the respective experiments are shown: five consecutive responses were averaged, and baseline levels are indicated by the dashed line and the pairing procedure by a red bar. Below the electrophysiological data, icons are shown at the times at which image stacks were acquired. The shape of the icon indicates whether new spines were detected. Scale bars for the morphological data: black, $10\ \mu\text{m}$; white, $3\ \mu\text{m}$. For animated time-lapse data of **d**, **e** and **g**, see Supplementary Information.

contain a postsynaptic density^{6,17}. It is therefore likely that, even if it is not present from the very beginning, ultimately a synapse will occupy a new spine. In any case, the first increase in synaptic strength cannot originate from newly grown spines, because these appear no earlier than 30 min after LTP induction. One possibility is that the initial physiological enhancement of synapses is only transient, and that it is later gradually replaced by the formation of new synaptic sites¹⁸ which then permanently contribute to the strengthening of overall synaptic weight.

We found that the disappearance of spines is not, as is the case in LTP and spine growth, controlled in a specific and activity-dependent manner, but that it rather occurs more-or-less randomly in time and space. Theoretically, there must also be a negative mechanism for the regulation of spine density if spines are not to increase constantly in number over the lifetime of a neuron. Whether the disappearance of spines, and thus synaptic weakening, occurs in a specific activity-dependent manner or by general synaptic decay has been a matter of debate^{19–21}. Our data support the latter, although we cannot rule out that other components, such as functional (rather than morphological) changes, are controlled in

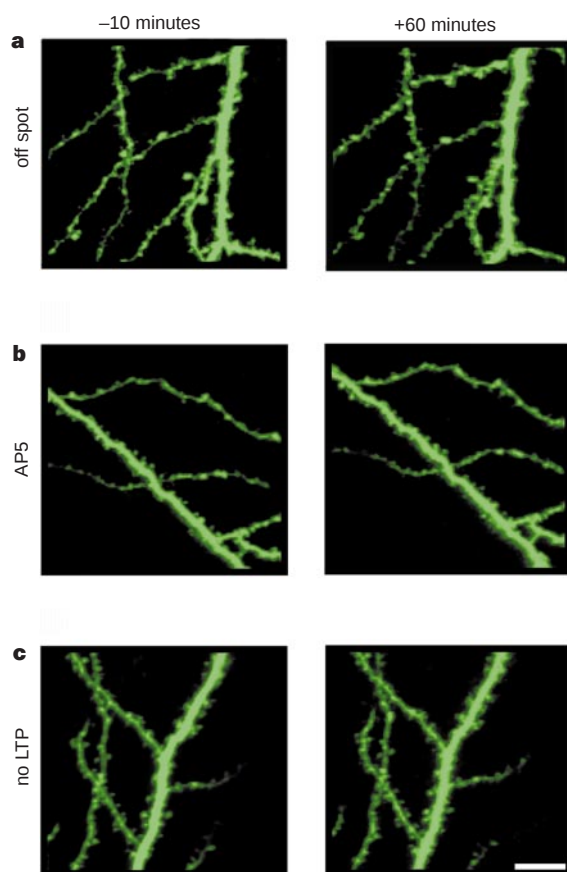


Figure 4. Control experiments. Three representative experiments with no change in synaptic efficacy. **a**, Images from a cell in which LTP was induced successfully, but the area shown is located well outside the superfused region ($>150\ \mu\text{m}$). **b**, Data from one of the experiments in which LTP induction was blocked by AP5. **c**, Data from an experiment in which the induction of LTP was unsuccessful. The remaining controls are summarized in Fig. 5 and Table 1. Scale bar: $10\ \mu\text{m}$.

an activity-dependent way.

In summary, we have shown that the induction of LTP in the hippocampus results in the emergence of new spines. The most attractive explanation for this phenomenon is the formation of new synapses that appear within an hour of the initiating stimulus. Although more experiments are necessary on the precise nature of

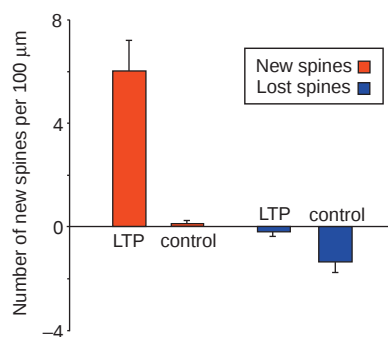


Figure 5 Quantitative results. The appearance and disappearance of spines was normalized to the complete length of examined dendrite in each individual neuron and the pooled results are plotted as the number of spines per $100\ \mu\text{m}$ of dendrite. For the calculation of the results for the control group, data were pooled from experiments in which LTP was blocked by AP5, LTP induction had been unsuccessful, or regions remote from the superfusion spot were chosen (Table 1). Error bars, standard errors of the mean.

these changes, our results show that not only physiological but also structural changes are important when neurons change the efficacy of their connections.

Note added in proof: While this paper was in press, evidence for growth of new dendritic processes induced by strong synaptic activity was reported (M. Maletic-Savatic, R. Malinow and K. Svoboda, *Science* **283**, 1923–1927; 1999). Although addressing a slightly different question, the results are in line with our observations reported here. □

Methods

Preparation of slice cultures. Organotypical slice cultures of rat hippocampus were prepared as described^{22,23} and selected for recording after 2–4 weeks in culture. Slice cultures consisting of only one or a few cell layers were used because the penetration depth of the superfusion technique is not much more than $\sim 30\ \mu\text{m}$ and therefore impractical in acute brain slices.

Data acquisition. Intracellular recordings were performed using an AxoClamp-2B amplifier (Axon Instruments) in current-clamp configuration. Sharp electrodes were used to prevent ‘run-down’ of cells due to washout of intracellular factors. Data were digitally sampled at 5 kHz, low-pass filtered at 3 kHz, and stored and analysed with custom made LabView software (National Instruments) on a PC. In all electrophysiology figures the amplitude of the EPSP (in mV) is plotted over time. The maximal slope of EPSPs was also determined in all experiments and the results are virtually the same apart from an expected slightly increased scatter of the data points.

Local superfusion and electrophysiology. Slice cultures were placed in the recording chamber and a stimulating electrode was positioned in the Schaffer collaterals. We impaled a pyramidal cell in the CA1 region with a recording electrode containing 3M potassium acetate and the fluorescent molecule calcein. Only cells with resting potentials $<-65\ \text{mV}$ were used for further experimentation. After a stable response to Schaffer collateral activation was recorded, we bathed the slice culture in a solution containing a moderate dose of Cd^{2+} ($10\ \mu\text{M}$) and a low concentration of Ca^{2+} ($0.8\ \text{mM}$) to prevent presynaptic Ca^{2+} influx, thereby abolishing transmitter release and, consequently, all postsynaptic responses. A local superfusion system^{2,3} containing normal extracellular solution (with a slightly elevated Ca^{2+} concentration³ of $5\ \text{mM}$) was then employed to restore locally synaptic responses in a region with a diameter of $\sim 30\ \mu\text{m}$. Using this technique, we ‘searched’ for synapses mediating responses from the stimulated Schaffer collaterals to the postsynaptically recorded neuron. After uncovering such a group of synapses, we recorded a baseline of the postsynaptic intracellular responses and acquired a first two-photon image of the filled neuron. After 10 min of baseline recording ($0.1\ \text{Hz}$ stimulation), synapses were enhanced by a pairing procedure (30 presynaptic stimuli concurrent with $1\ \text{nA}$ postsynaptic current injections at an interstimulus interval of $10\ \text{s}$). Subsequently, EPSP amplitudes were recorded for up to 2 h until the electrode was withdrawn. After the removal of the electrode, the neuron could be observed for up to 28 h with the two-photon microscope.

Two-photon microscopy. A scanhead for confocal microscopy (Biorad 1024, Biorad) was modified for two-photon microscopy and equipped with a new photomultiplier (Hamamatsu Photonics) for external detection of fluorescent light. A ‘Tsunami Lite’ titanium-sapphire femtosecond laser pumped by a ‘Millennia’ neodymium yttrium vanadate solid-state laser (both SpectraPhysics) served as the two-photon light source. The scanhead was connected to a Zeiss Axiovert microscope and images were acquired with a resolution of $0.33\ \mu\text{m}$ in the x and y directions and $1\text{--}1.8\ \mu\text{m}$ in the z direction using a $63\times$ Neofluar, oil-immersion objective. They were averaged between one and three times depending on the concentration of the fluorescent dye. The relatively large spacing in the z -direction was needed to achieve a time resolution of roughly one image per 10 min. Closer spacing would have resulted in too much laser power per unit time, causing noticeable photodynamic damage. It is conceivable that, because of the spacing, we missed the smallest spines, which might explain why the overall spine density was lower than that reported in the literature⁸.

For volume rendering, we used the Imaris (BitPlane) raytrace-shadow projection algorithm on an SGI O2 (Silicon Graphics). No further filtering or deconvolution algorithms were applied.

The online 3D imaging of the neuron greatly facilitated the localization of

dendritic branches suitable for superfusion because it allowed us to select those synapses that were located close to the surface of the slice culture (Fig. 1c, right). When the dendrites were located more than 30 μm below the surface, no postsynaptic EPSP could be elicited ($n = 85$ experiments), indicating that this is the maximal penetration depth of the superficially applied superfusion spot. Thus for further analysis, only parts of the dendrite closer than 30 μm to the surface were considered to be potential sites of active synapses.

Blind evaluation. The data were blindly evaluated by an observer who was not familiar with the experiments and not directly involved in the study. The observer was confronted with pairs of data sets (before and 1 h after pairing). These data sets included: (1) all experiments in which pairing had successfully induced LTP (7 experiments); (2) those experiments in which pairing was applied but LTP was prevented by 50 μM AP5 (3 experiments); (3) three pairs of data sets that came from successful experiments but were from parts of the dendrite where activity was blocked and where therefore no changes were expected; and (4) two pairs of data sets in which the pairing procedure did not result in a long-lasting potentiation. The observer received the original data sets (stacks of images) and carefully compared spines before and after pairing by screening through the individual 'sections' using the 'Confocal Assistant' software (Biorad). All spines were labelled and given a rating between one (barely visible) and three (head and neck clearly distinguishable). After complete evaluation, the code was broken and (2), (3) and (4) were pooled into a 'control' group. To reduce the noise in this procedure, only spines with a difference in the above rating greater than one ($\Delta > 1$) between the two time frames were considered to be appearing or vanishing spines. They were entered into the statistics and plotted as red (or blue) dots on a picture of the dendritic tree which was overlaid with the superfusion spot (Fig. 3). To correct for the length of dendrite examined in each experiment, a weighted average was used to calculate the means and t -values. We also verified that including smaller changes in spine morphology ($\Delta = 1$) in the analysis did not change the basic results.

Received 22 February; accepted 7 April 1999.

- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Veslovsky, N. S., Engert, F. & Lux, H. D. Fast local superfusion technique. *Pflügers Arch.* **432**, 351–354 (1996).
- Engert, F. & Bonhoeffer, T. Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388**, 279–284 (1997).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
- Woolley, C. S., Gould, E., Frankfurt, M. & McEwen, B. S. Naturally occurring fluctuation in dendritic spine density on adult hippocampal pyramidal neurons. *J. Neurosci.* **10**, 4035–4039 (1990).
- Moser, M.-B., Trommald, M. & Andersen, P. An increase in dendritic spine density on hippocampal CA1 pyramidal cells following spatial learning in adult rats suggests the formation of new synapses. *Proc. Natl Acad. Sci. USA* **91**, 12673–12675 (1994).
- Collin, C., Miyaguchi, K. & Segal, M. Dendritic spine density and LTP induction in cultured hippocampal slices. *J. Neurophysiol.* **77**, 1614–1623 (1997).
- McKinney, R. A., Capogna, M., Dürr, R., Gähwiler, B. H. & Thompson, S. M. Miniature synaptic events maintain dendritic spines via AMPA receptor activation. *Nature Neurosci.* **2**, 44–49 (1999).
- Desmond, N. L. & Levy, W. B. Morphological correlates of long-term potentiation imply the modification of existing synapses, not synaptogenesis, in the hippocampal dentate gyrus. *Synapse* **5**, 139–143 (1990).
- Hosokawa, T., Rusakov, D. A., Bliss, T. V. P. & Fine, A. Repeated confocal imaging of individual dendritic spines in the living hippocampal slice: Evidence for changes in length and orientation associated with chemically induced LTP. *J. Neurosci.* **15**, 5560–5573 (1995).
- Buchs, P. A. & Müller, D. Induction of long-term potentiation is associated with major ultrastructural changes of activated synapses. *Proc. Natl Acad. Sci. USA* **93**, 8040–8045 (1996).
- Sorra, K. E. & Harris, K. M. Stability in synapse number and size at 2 hr after long-term potentiation in hippocampal area CA1. *J. Neurosci.* **18**, 658–671 (1998).
- Crick, F. Do dendritic spines twitch? *Trends Neurosci.* **5**, 44–46 (1982).
- Svoboda, K., Tank, D. W. & Denk, W. Direct measurement of coupling between dendritic spines and shafts. *Science* **272**, 716–719 (1996).
- Fischer, M., Kaech, S., Knutti, D. & Matus, A. Rapid actin-based plasticity in dendritic spines. *Neuron* **20**, 847–854 (1998).
- Dunaevsky, A., Heintz, N., Mason, C. A. & Yuste, R. Two-photon imaging of dendritic spines of cerebellar and cortical cells transfected with GFP. *Soc. Neurosci. Abstr.* **24**, 316.1 (1998).
- Gray, E. G. Electron microscopy of synaptic contacts on dendritic spines of the cerebral cortex. *Nature* **183**, 1592–1593 (1959).
- Bolshakov, V. Y., Golan, H., Kandel, E. R. & Siegelbaum, S. A. Recruitment of new sites of synaptic transmission during the cAMP-dependent late phase of LTP at CA3–CA1 synapses in the hippocampus. *Neuron* **19**, 635–651 (1997).
- Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
- Miller, K. D. & MacKay, D. C. The role of constraints in Hebbian learning. *Neural Comput.* **6**, 100–126 (1994).
- Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896 (1998).
- Gähwiler, B. H. Organotypic monolayer cultures of nervous tissue. *J. Neurosci. Meth.* **4**, 329–342 (1981).
- Bonhoeffer, T., Staiger, V. & Aertsen, A. Synaptic plasticity in rat hippocampal slice cultures: Local 'Hebbian' conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement. *Proc. Natl Acad. Sci. USA* **86**, 8113–8117 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements. We thank I. Kehr for help with the evaluation of the data, and M. Hübener, M. Korte and M. Meister for comments on the manuscript.

Correspondence and requests for materials should be addressed to T.B.

A family of mammalian Na⁺-dependent L-ascorbic acid transporters

Hiroyasu Tsukaguchi*†, Taro Tokui*†‡, Bryan Mackenzie*†, Urs V. Berger*, Xing-Zhen Chen*, Yangxi Wang*, Richard F. Brubaker§ & Matthias A. Hediger*||

* Membrane Biology Program and Renal Division, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, and

|| Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA

§ Department of Ophthalmology, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55905, USA

† These authors contributed equally to this work.

Vitamin C (L-ascorbic acid) is essential for many enzymatic reactions, in which it serves to maintain prosthetic metal ions in their reduced forms (for example, Fe²⁺, Cu⁺)^{1,2}, and for scavenging free radicals in order to protect tissues from oxidative damage³. The facilitative sugar transporters of the GLUT type can transport the oxidized form of the vitamin, dehydroascorbic acid^{4–6}, but these transporters are unlikely to allow significant physiological amounts of vitamin C to be taken up in the presence of normal glucose concentrations, because the vitamin is present in plasma essentially only in its reduced form⁷. Here we describe the isolation of two L-ascorbic acid transporters, SVCT1 and SVCT2, from rat complementary DNA libraries, as the first step in investigating the importance of L-ascorbic acid transport in regulating the supply and metabolism of vitamin C. We find that SVCT1 and SVCT2 each mediate concentrative, high-affinity L-ascorbic acid transport that is stereospecific and is driven by the Na⁺ electrochemical gradient. Despite their close sequence homology and similar functions, the two isoforms of the transporter are discretely distributed: SVCT1 is mainly confined to epithelial systems (intestine, kidney, liver), whereas SVCT2 serves a host of metabolically active cells and specialized tissues in the brain, eye and other organs.

We isolated a 2,472-base-pair complementary DNA coding for a 604-amino-acid protein, SVCT1 (sodium-dependent vitamin C transporter 1), by screening a rat kidney cDNA library for Na⁺-dependent L-[¹⁴C]ascorbic acid transport activity in RNA-injected *Xenopus* oocytes (Fig. 1). Subsequent polymerase chain reaction (PCR)-based homology screening yielded a related cDNA (of 6.5 kilobases (kb), from rat brain) coding for a 592-amino-acid protein, SVCT2, with 65% amino-acid identity to SVCT1. SVCT1 and SVCT2 have similar hydropathy profiles, each predicting 12 putative membrane-spanning domains. SVCT2 is 84% identical (at the amino-acid level) to a predicted partial polypeptide encoded by the human KIAA0238 gene⁸; therefore, KIAA0238 may be the species homologue of rat SVCT1. SVCT1 is 59% identical to KIAA0238 and 32% identical to a mouse yolk-sac permease-like molecule (mYspl1)⁹ of unknown function. SVCT1 and SVCT2 share weak homology with putative permeases of *Caenorhabditis elegans* and *Arabidopsis thaliana*, and with a bacterial purine/pyrimidine permease¹⁰. We found no significant sequence homology between SVCT and any other known family of mammalian membrane transporters¹¹; hence SVCT1 and

‡ Present address: Sankyo Company Ltd, Analytical and Metabolic Research Laboratories, 2-58 Hiromachi 1-Chome, Shinagawa-ku, Tokyo 140, Japan.

SVCT2 constitute a novel mammalian transporter family with a potential evolutionary link to purine/pyrimidine transporters found in prokaryotes and eukaryotes.

We investigated the functional characteristics of the SVCT protein family by expressing SVCT1 and SVCT2 in *Xenopus* oocytes and applying radiotracer and voltage-clamp techniques. Our data indi-

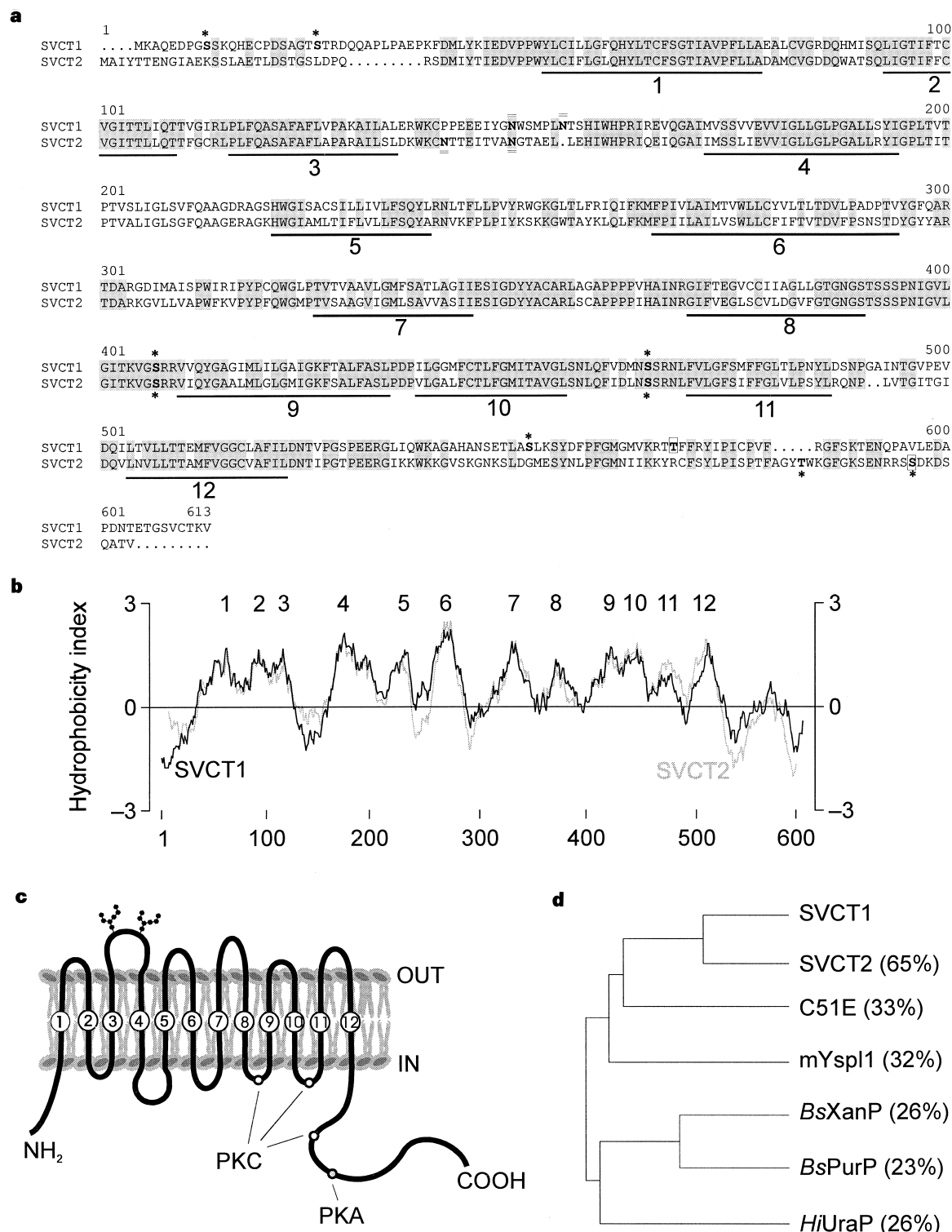


Figure 1 Sequence and putative membrane topology of SVCT1 and SVCT2. **a**, Amino-acid sequence alignment of SVCT1 with SVCT2. Putative membrane-spanning regions are underlined and numbered 1–12. Conserved residues are highlighted by shading. Potential sites for *N*-linked glycosylation (Asn 145, 151 for SVCT1 and Asn 132, 140 for SVCT2) (double dash), protein kinase C phosphorylation (Ser 9, 22, 403, 461, 547 for SVCT1 and Ser 397, 455, 584, Thr 571 for SVCT2) (asterisk) and cAMP-dependent phosphorylation (Thr 564 for SVCT1 and Ser 584 for SVCT2) (box). **b**, Kyte–Doolittle hydrophobicity plots of SVCT1 (black) and

SVCT2 (grey) were generated using a 21-amino-acid window, with numbered putative membrane-spanning domains. **c**, Structural model of SVCT1 based on its hydropathy profile. **d**, SVCT superfamily phylogenetic tree, with related amino-acid sequences aligned against SVCT1: C51E, *Caenorhabditis elegans* C51E3.6 (GenBank accession number Z78410); mYsp1, mouse yolk sac permease-like protein (U25739); BsXanP, *Bacillus subtilis* xanthine permease (P42086); BsPurP, *B. subtilis* uric acid permease (Z99120); HiUraP, *Haemophilus influenzae* uracil permease (P45117).

cate that SVCT1 and SVCT2 each mediate saturable, Na^+ -dependent L-ascorbic acid transport in a rheogenic manner. L-[^{14}C]ascorbic acid uptake (in 100 mM NaCl) followed Michaelis–Menten-type saturation kinetics with high apparent affinity for L-ascorbic acid (see Table 1 in Supplementary Information). Rates of uptake were not affected by the replacement of chloride with gluconate (data not shown), indicating that neither isoform was Cl^- -dependent.

In the presence of Na^+ , L-ascorbic acid evoked reversible inward currents of up to -100 nA in oocytes expressing SVCT1, and up to -10 nA in oocytes expressing SVCT2 (Fig. 2). Evoked currents in control oocytes were -1 ± 0.5 nA (s.d.) at $500 \mu\text{M}$ L-

ascorbic acid. Transporter-associated currents were investigated in detail for SVCT1 and, where possible, our findings were supported by data from radiotracer experiments for SVCT1 and SVCT2. In oocytes expressing SVCT1, the currents evoked by L-ascorbic acid (at saturating levels) showed a curvilinear dependence on membrane potential (V_m) (Fig. 2b), saturating with hyperpolarization (-70 mV). The current/voltage relationship was roughly linear between -50 and $+30$ mV (tending towards a zero-current asymptote at positive V_m), with no reversal of the currents observed up to $+50$ mV. At -50 mV, $K_{0.5}^{\text{Asc}}$ (the L-ascorbic acid concentration at which currents were half-maximal) was $30 \mu\text{M}$ (Fig. 2c) with a

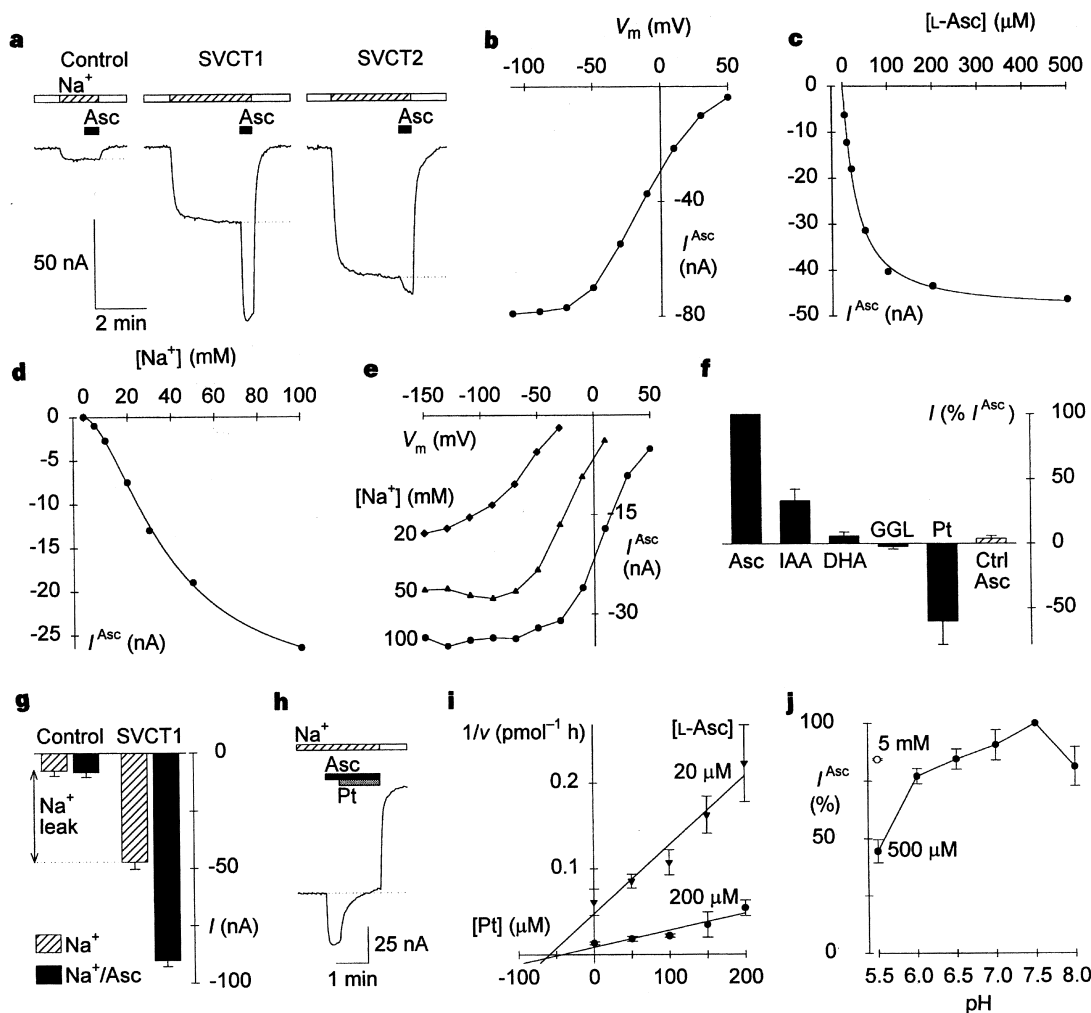


Figure 2 Functional characterization of SVCT expressed in *Xenopus* oocytes. **a**, Typical current recordings from a control oocyte, and oocytes expressing SVCT1 and SVCT2, all from a single batch of oocytes. The approximate baseline current in standard Na^+ medium at pH 7.5 (hatched boxes) is indicated by the dotted line. $500 \mu\text{M}$ L-ascorbic acid (Asc) was superfused (in Na^+ medium) for the periods indicated by the black boxes, and substrates were always washed out in choline medium (blank boxes). **b**, Current/voltage (I/V) relationship for SVCT1 at $500 \mu\text{M}$ (saturating) L-ascorbic acid (at 100 mM Na^+). **c**, L-Ascorbic acid saturation kinetics for SVCT1 at 100 mM Na^+ and at -50 mV ($K_{0.5}^{\text{Asc}} = 29 \pm 3 \mu\text{M}$; $I_{\text{max}}^{\text{Asc}} = -49 \pm 1$ nA; $n_H = 1.1 \pm 0.1$). **d**, Na^+ saturation kinetics of the currents evoked by $200 \mu\text{M}$ L-ascorbic acid for SVCT1 at -50 mV (n_H for Na^+ = 1.7 ± 0.1 ; $K_{0.5}^{\text{Na}} = 38 \pm 2$ mM; $I_{\text{max}}^{\text{Na}} = -31 \pm 1$ nA). **e**, The effect of Na^+ on the I/V relationship for SVCT1 (at $200 \mu\text{M}$ L-ascorbic acid). **f**, Substrate selectivity of SVCT1. Currents (mean $\pm$ s.e.m., $n = 7$ –12 oocytes) evoked by test substrates applied at $500 \mu\text{M}$ (at -50 mV) were normalized to the L-ascorbic acid-evoked current (-25 ± 4 nA, $n = 12$); IAA, D-isoascorbic acid; DHA, dehydroascorbic acid; GGL, L-gulonolactone; Pt, phloretin; Ctrl Asc, L-ascorbic acid-evoked current in control oocytes. **g**, Comparison of currents due to Na^+ leak (uniport) and Na^+ /L-ascorbic acid (Na^+ /

Asc) co-transport pathways in SVCT1. Currents resulting from the addition of either 100 mM Na^+ alone (hatched bars) or 100 mM Na^+ plus $500 \mu\text{M}$ L-ascorbic acid (Na^+ /Asc) (solid bars) were compared in control oocytes and oocytes expressing SVCT1 (mean $\pm$ s.e.m., $n = 5$). Zero represents the current in Na^+ -free medium. **h**, Phloretin (Pt) inhibition of the L-ascorbic acid-evoked current in SVCT1. An oocyte expressing SVCT1 was superfused (at -50 mV) in standard Na^+ medium (hatched box) before the addition of $200 \mu\text{M}$ L-ascorbic acid (black box) and $100 \mu\text{M}$ phloretin (grey box), before washing out with choline medium (blank box); the approximate baseline current in Na^+ is indicated by the dotted line. **i**, Dixon analysis of phloretin inhibition. L-[^{14}C]ascorbic acid uptake, v (mean $\pm$ s.e.m. from 6–10 oocytes, 60 min), was determined at $20 \mu\text{M}$ or $200 \mu\text{M}$ L-ascorbic acid in the presence of 0–200 μM phloretin. From linear regression, the reciprocal plots ($20 \mu\text{M}$, $r^2 = 0.90$; $200 \mu\text{M}$, $r^2 = 0.95$) intersected at $\sim 65 \mu\text{M}$ ($-K_{0.5}^{\text{Pt}}$). **j**, pH dependence of L-ascorbic acid-evoked currents in SVCT1. Currents evoked by $500 \mu\text{M}$ L-ascorbic acid (filled circles) at extracellular pH 5.5–8.0 were normalized to the current at pH 7.5 (-28 ± 2 nA, mean $\pm$ s.e.m., $n = 6$). The current evoked by 5 mM L-ascorbic acid at pH 5.5 is also shown (open circle).

Hill coefficient (n_H) of ~ 1 . The n_H for Na^+ was ~ 2 (Fig. 2d). $K_{0.5}^{\text{Na}}$ (at $200 \mu\text{M}$ L-ascorbic acid) was 40 mM at -50 mV ; $K_{0.5}^{\text{Na}}$ was not significantly different at hyperpolarized V_m , but rose at depolarized V_m (data not shown). These data indicate that two Na^+ ions bind to SVCT1 and that binding is voltage-sensitive. SVCT1-mediated L-ascorbic acid transport is driven by the electrochemical gradient for Na^+ because, at any given V_m , the current evoked by $200 \mu\text{M}$ L-ascorbic acid was greatest at higher Na^+ concentrations (Fig. 2e). The evoked currents obtained at 100 mM Na^+ saturated at a less negative V_m and with a larger maximal current than those obtained at 50 and 20 mM Na^+ . Following step-changes in voltage (not shown), SVCT1 exhibited pre-steady-state currents (decaying with time constants of 10 – 40 ms). These pre-steady-state currents were sensitive to changes in extracellular Na^+ concentration and (in analogy to other ion-coupled transporters^{12,13}) are due in part to binding/dissociation of the driving ion (in this case Na^+) to the transporter. Switching from 0 to 100 mM Na^+ in the absence of L-

ascorbic acid caused a small inward current in control oocytes (Fig. 2a, g), but a much larger current in oocytes expressing either SVCT1 or SVCT2. Addition of L-ascorbic acid in the presence of Na^+ evoked still larger inward currents. These observations indicate that SVCT1 and SVCT2 can additionally exhibit a Na^+ leak (uniport) current (similar to the leak pathways in other ion-coupled transporters^{14–16}) that is as much as half the magnitude of the current associated with Na^+ /L-ascorbic acid co-transport (Fig. 2g).

SVCT1 was highly selective for L-ascorbic acid, which evoked much larger currents than did D-isoascorbic acid ($\sim 30\%$ of that for L-ascorbic acid) or dehydroascorbic acid ($\sim 5\%$) (Fig. 2f). (Dehydroascorbic acid did not evoke a current in control oocytes.) D-glucose, uracil (not shown) and intermediates of vitamin C metabolism (such as L-gulono- γ -lactone) were excluded. Several test compounds, including aspirin (acetylsalicylic acid), xanthine, sulfinpyrazone and phlorizin, each evoked tiny outward currents in oocytes expressing SVCT1 and also slightly inhibited SVCT1- or

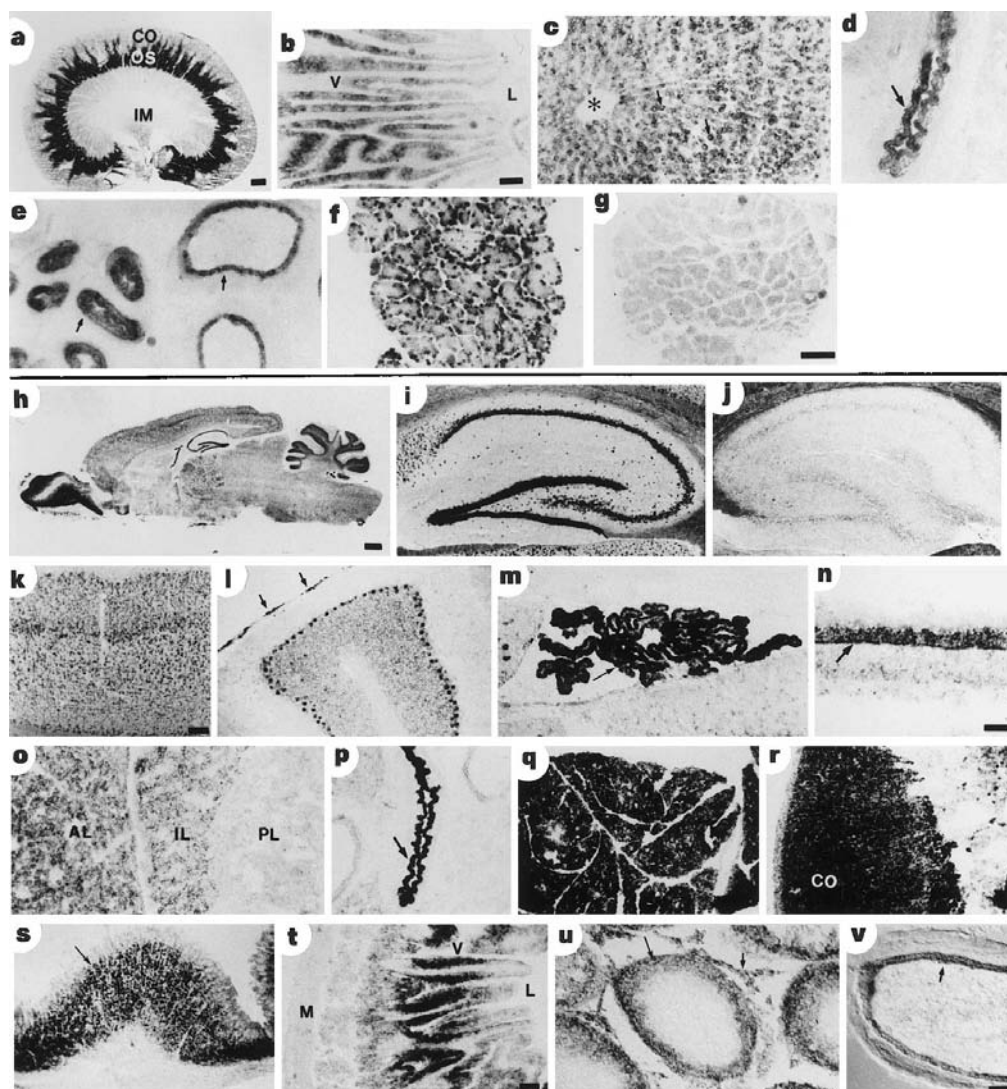


Figure 3 Localization of SVCT1 and SVCT2 mRNA in rat tissues detected by *in situ* hybridization. Bright-field micrographs of cryosections hybridized to digoxigenin-labelled antisense cRNA probes of SVCT1 (a–f) and SVCT2 (h, i, k–v). Those hybridized to sense cRNA probes for SVCT1 (g) and SVCT2 (j) revealed no significant signal. **a**, Kidney (CO, cortex; OS, outer stripe of the medulla; IM, inner medulla). **b**, Duodenum (V, villi; L, lumen). **c**, Liver (asterisk, central vein; arrows indicate hepatocytes). **d**, Lung (arrow, bronchiolar epithelium). **e**, Epididymis (arrows, tubules). **f**, Lacrimal gland. **g**, Lacrimal gland (sense). **h**, Brain. **i**, Hippocampus. **j**, Hippocampus (sense). **k**, Cerebral cortex. **l**, Cerebellum (arrows,

labelled meningeal layers). **m**, Choroid plexus (arrow). **n**, Retina (arrow, inner nuclear layer). **o**, Pituitary gland (AL, anterior lobe; IL, intermediate lobe; PL, posterior lobe). **p**, Lung (arrow, bronchiolar epithelium). **q**, Pancreas. **r**, Adrenal gland (CO, cortex). **s**, Stomach (arrow, gastric glands). **t**, Duodenum (M, muscle layer; V, villi; L, lumen). **u**, Testis (long arrow, spermatocytes; short arrow, interstitial cells). **v**, Epididymis (arrow, tubular epithelium). Scale bars: **a**, **h**, 1 mm ; **b**, $100 \mu\text{m}$; **c**–**g**, $100 \mu\text{m}$ (shown in **g**); **i**, **p**–**t**, $100 \mu\text{m}$ (shown in **t**); **i**–**k**, $200 \mu\text{m}$ (shown in **k**); **m**, **o**, **u**, **v**, $50 \mu\text{m}$ (shown in **v**); **n**, $50 \mu\text{m}$.

SVCT2-mediated L-[14 C]ascorbic acid uptake (data not shown); that is, they exhibited characteristics of weak blockers. Phloretin evoked a sizeable outward current in oocytes expressing SVCT1 (Fig. 2f). We attribute this effect to phloretin blocking the Na^+ leak current. In addition, phloretin blocked the inward current evoked by L-ascorbic acid (Fig. 2h) and inhibited L-[14 C]ascorbic acid uptake by SVCT1 or SVCT2. Dixon analysis of L-[14 C]ascorbic acid uptake for SVCT1 (Fig. 2i) revealed that phloretin was a non-competitive inhibitor¹⁷ with $K_i \approx 65 \mu\text{M}$. Phloretin also inhibited the Na^+ leak current with an apparent $K_i \approx 100 \mu\text{M}$ (data not shown). The similarity of these values for the effects of phloretin on both the co-transport and uniport modes of the transport cycle indicates that phloretin may interact with SVCT1 at a single locus on the protein. The evoked currents in SVCT1 were sensitive to changes in extracellular pH (Fig. 2j). Those at pH 5.5 were 50% smaller than those at pH 7.5, but were partly restored by the addition of excess L-ascorbic acid (5 mM). SVCT2 showed a similar pattern of pH sensitivity in radiotracer experiments (not shown). This pH sensitivity is probably a result of reduced binding affinities for L-ascorbic acid, rather than less available L-ascorbic acid in the deprotonated (1^-) form, as more than 95% is in the 1^- form at pH 5.5 ($\text{p}K_{\text{a}1} = 4.2$).

The function of SVCT2 was primarily investigated by radio tracer uptake studies because the currents were relatively small. The studies did not reveal any functional differences between the two SVCT isoforms. The activities of SVCT1 and SVCT2 expressed in oocytes were consistent with those of several mammalian tissues (in vesicles, isolated tissues or cultured cell lines)^{18–22} with regard to the following characteristics: high apparent affinity for L-ascorbic acid ($K_{0.5}^{\text{asc}}$ of 10–100 μM); a preference for L-ascorbic acid over D-isoascorbic acid, which is preferred over dehydroascorbic acid; and Na^+ dependence.

We found striking differences between SVCT1 and SVCT2 in terms of tissue distribution. Northern blot analysis using a probe for SVCT1 revealed intense bands at $\sim 2.5 \text{ kb}$ and $\sim 4.0 \text{ kb}$ in kidney, intestine and liver, whereas probing for SVCT2 RNA resulted in a weaker signal at $\sim 6.5 \text{ kb}$ for those tissues (see Fig. 5 in Supplementary Information). Using *in situ* hybridization (Fig. 3), SVCT1 was localized to the straight segment (S3) of the proximal tubule in the kidney, consistent with studies of rat kidney L-ascorbic acid transport *in vitro*^{18,19}. SVCT1 and SVCT2 messenger RNAs were present in enterocytes in the small intestine. In liver, SVCT1 but not SVCT2 was detected in hepatocytes. Both isoforms were detected in epithelial cells of the bronchiole and epididymis.

SVCT2 was abundantly expressed in a range of neural, neuroendocrine, exocrine and endothelial tissues as well as in osteoblasts. Northern blot analysis gave a strong SVCT2 signal for brain at $\sim 6.5 \text{ kb}$. The mRNA was localized to neurons throughout the central nervous system. The distribution closely matched the sites of rapid neuronal accumulation observed in mouse brain following intravenous injection of L-[14 C]ascorbic acid²³ and is consistent with the view that neurons take up L-ascorbic acid (released by astroglia). SVCT2 was also detected in the meninges and choroid plexus, indicating that SVCT2 may facilitate entry of L-ascorbic acid into the cerebrospinal fluid compartments. In the retina, SVCT2 labelling was observed exclusively in the inner nuclear layer (the site of bipolar, amacrine and horizontal-cell bodies).

L-ascorbic acid is believed to be crucial in protecting components of the eye (such as the lens and cornea) from radiation-induced damage²⁴. As the levels of L-ascorbic acid in the aqueous humour (anterior chamber) in diurnal mammals can be around 20-fold greater than in nocturnal mammals²⁵, we investigated whether the distribution of SVCT isoforms in the eye differed between the rat (nocturnal) and the rabbit (diurnal–nocturnal). In the rat, neither SVCT1 nor SVCT2 was detected in the ciliary body (which secretes aqueous humour). However, in the albino rabbit, SVCT2 was abundantly expressed in the pigmented epithelium of the ciliary

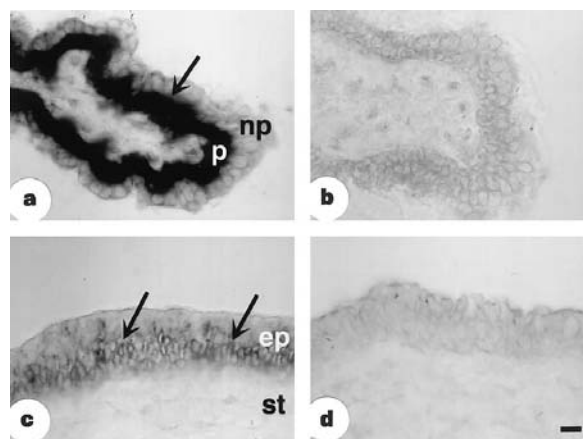


Figure 4 Localization of SVCT2 mRNA in the albino rabbit eye detected by *in situ* hybridization. Bright-field micrographs of cryosections hybridized to digoxigenin-labelled antisense cRNA probe showing: **a**, the pigmented (p) and nonpigmented (np) cell layers in ciliary body and ciliary processes; **b**, sense control for ciliary process; **c**, Deep layers of the corneal epithelium (arrows); ep, epithelium; st, stroma of the cornea; **d**, Sense control for corneal epithelium. Scale bar, 20 μm .

body, and moderately expressed in the deeper layers of the corneal epithelium (Fig. 4). This SVCT2 distribution was consistent with the known sites of L-ascorbic acid concentration in the rabbit eye²⁶, and the characteristics of SVCT2 reflected those of a Na^+ -dependent L-ascorbic acid transport activity expressed in bovine pigmented ciliary epithelial cells²⁰. In the rat, SVCT1 (Fig. 3) and SVCT2 (not shown) were each detected in exocrine cells of the lacrimal gland. There, the supply of L-ascorbic acid to the tear fluid may provide sufficient antioxidant protection in nocturnal species.

SVCT2 was expressed in several components of the endocrine system, including the anterior pituitary, in which L-ascorbic acid is important in α -amidation of peptide hormones. Labelling was also observed in the intermediate lobe but not in the posterior lobe. SVCT2 was abundantly expressed throughout the pancreas and in the adrenal cortex. In the adrenal gland, L-ascorbic acid is critical for the copper-associated dopamine- β -hydroxylase involved in norepinephrine synthesis². SVCT2 was detected in gastric glands, extending from the base to the isthmus. Autoradiographic studies in mouse demonstrated rapid accumulation of L-ascorbic acid into the gastric mucosa following intravenous injection²³. These observations implicate SVCT2 in the basolateral uptake of L-ascorbic acid, for secretion by the gastric gland, aiding dietary absorption of iron. An intense SVCT2-related signal was obtained in northern blot analysis of poly(A)⁺ RNA from the murine osteoblast MC3T3-E1 cell line²². SVCT2-mediated L-ascorbic acid transport into osteoblasts may account for the adequate supply of the vitamin to maintain Fe^{2+} required by hydroxylases involved in collagen synthesis². Among immune system organs, SVCT2 was detected by northern blot in the spleen and thymus. In testis, SVCT2 was expressed in interstitial cells and spermatocytes.

To assess the relative contributions to L-ascorbic acid transport of SVCT1 and SVCT2 or possible additional transporters, we performed hybrid depletion of poly(A)⁺ RNA isolated from several rat tissues and monitored transport activity in oocytes (see Fig. 6 in Supplementary Information). L-[14 C]ascorbic acid uptake was threefold higher in oocytes injected with intestinal RNA than in control oocytes. The induced activity was Na^+ -dependent (not shown) and abolished by hybridization with antisense oligonucleotide against SVCT1, indicating that SVCT1 is the predominant L-ascorbic acid transport system in the small intestine. Similar results were obtained for kidney, and data for adrenal gland indicate that expression of SVCT2 and SVCT1 can account for the increased uptake in RNA-injected oocytes. SVCT1 and SVCT2 therefore seem

to be the predominant L-ascorbic acid transport systems (at least for $K_{0.5}^{asc}$ of similar order) in the tissues tested.

Our data show that SVCT1 and SVCT2 play central roles in the absorption and accumulation of vitamin C in many tissues. SVCT1 is largely confined to bulk-transporting epithelia, in which it may serve whole-body homeostasis and metabolic requirements for the vitamin, whereas SVCT2 may account for the widespread tissue-specific uptake of L-ascorbic acid required for an array of biological functions, including enzymatic reactions and antioxidation. To serve these roles, we speculate that SVCT1 is located on the apical membrane, whereas SVCT2 could be directed to the basolateral membrane in endocrine or exocrine tissues. However, such polarization must be verified by immunocytochemistry. Despite their discrete distribution, SVCT1 and SVCT2 displayed very similar functional characteristics when expressed in oocytes. Molecular cloning of SVCT1 and SVCT2 now enables us to examine the possibility of differential regulation of these two transporters and how they may adapt to oxidative stress in pathological conditions. *Note added in proof.* Related sequences have recently been reported but their function remains unknown (AF058319, AF058317, *Biochim. Biophys. Acta* **1442**, 353–360, 1998). □

Methods

Detailed methods are provided in the Supplementary Information. Briefly, the SVCT1 cDNA was isolated from rat kidney by expression cloning as described²⁷ and the SVCT2 cDNA was subsequently isolated by PCR-based homology screening of a rat brain cDNA library. A 1.2-kb rabbit cDNA with 96% identity to residues 96–505 of the rat SVCT2 was obtained by RT-PCR. Radiotracer and voltage-clamp experiments were performed (at 22°C) in *Xenopus* oocytes (isolated and maintained as described²⁸) 2–7 days after injection with ~25 ng of SVCT1 or SVCT2 cRNA synthesized *in vitro*. Radiotracer uptake was determined by incubating 6–10 oocytes for 30 or 60 min in standard 100 mM Na⁺ or Na⁺-free media with 10–600 μM L-[1-¹⁴C]ascorbic acid. Steady-state currents measured using a two-microelectrode voltage-clamp were fitted to equation 11.3 of ref. 28. Protocols used in northern blot analysis, *in situ* hybridization (with digoxigenin-labelling as described²⁹) and hybrid depletion studies are detailed in the Supplementary Information.

Received 8 February; accepted 17 March 1999.

- Englard, S. & Seifter, S. The biochemical functions of ascorbic acid. *Annu. Rev. Nutr.* **6**, 365–406 (1986).
- Padh, H. Vitamin C: newer insights into its biochemical functions. *Nutr. Rev.* **49**, 65–70 (1991).
- Rose, R. C. & Bode, A. M. Biology of the free radical scavengers: an evaluation of ascorbate. *FASEB J.* **7**, 1135–1142 (1993).
- Vera, J. C., Rivas, C. I. & Fischberg, J. Mammalian facilitative hexose transporters mediate the transport of dehydroascorbic acid. *Nature* **364**, 79–82 (1993).
- Rumsey, S. C. *et al.* Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid. *J. Biol. Chem.* **272**, 18982–18989 (1997).
- Agus, D. B. *et al.* Vitamin C crosses the blood-brain barrier in the oxidized form through the glucose transporters. *J. Clin. Invest.* **100**, 2842–2848 (1997).
- Dhariwal, K. R., Hartzell, W. O. & Levine, M. Ascorbic acid and dehydroascorbic acid measurements in human plasma and serum. *Am. J. Clin. Nutr.* **54**, 712–716 (1991).
- Nagase, T. *et al.* Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (K1AA0201–K1AA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res.* **3**, 321–354 (1996).
- Guimaraes, M. J. *et al.* A new approach to the study of haematopoietic development in the yolk sac and embryoid bodies. *Development* **121**, 3335–3346 (1995).
- Diallinas, G., Gorfinkiel, L., Arst, H. N., Cecchetto, G. & Scaccocchio, C. Genetic and molecular characterization of a gene encoding a wide specificity purine permease of *Aspergillus nidulans* reveals a novel family of transporters conserved in prokaryotes and eukaryotes. *J. Biol. Chem.* **270**, 8610–8622 (1995).
- Wright, E. M., Loo, D. D. F., Turk, E. & Hirayama, B. A. Sodium cotransporters. *Curr. Opin. Cell Biol.* **8**, 468–473 (1996).
- Hazama, A., Loo, D. D. F. & Wright, E. M. Presteady-state currents of the rabbit Na⁺/glucose cotransporter (SGLT1). *J. Membr. Biol.* **155**, 175–186 (1997).
- Mackenzie, B., Loo, D. D. F., Panayotova-Heiermann, M. & Wright, E. M. Biophysical characteristics of the pig kidney Na⁺/glucose cotransporter SGLT2 reveal a common mechanism for SGLT1 and SGLT2. *J. Biol. Chem.* **271**, 32678–32683 (1996).
- Mackenzie, B., Loo, D. D. F. & Wright, E. M. Relationships between Na⁺/glucose cotransporter currents and fluxes. *J. Membr. Biol.* **162**, 101–106 (1998).
- Gunshin, H. *et al.* Cloning and characterization of a proton-coupled mammalian metal-ion transporter. *Nature* **388**, 482–488 (1997).
- Eskandari, S. *et al.* Thyroid Na⁺/I⁻ symporter: mechanism, stoichiometry, and specificity. *J. Biol. Chem.* **272**, 27230–27238 (1997).
- Segel, I. H. *Biochemical Calculations* 2nd edn (Wiley, New York, 1976).
- Bowers-Komro, D. M. & McCormick, D. B. Characterization of ascorbic acid uptake by isolated rat kidney cells. *J. Nutr.* **121**, 57–64 (1991).
- Toggenburger, G. *et al.* Na⁺-dependent, potential-sensitive L-ascorbate transport across brush border membrane vesicles from kidney cortex. *Biochim. Biophys. Acta* **646**, 433–443 (1981).

- Helbig, H. *et al.* Electrogenic Na⁺-ascorbate cotransport in cultured bovine pigmented ciliary epithelial cells. *Am. J. Physiol.* **256**, C44–C49 (1989).
- Rose, R. C. & Wilson, J. X. in *Vitamin C in Health and Disease* (eds Packer, L. & Fuchs, J.) 143–161 (Dekker, New York, 1997).
- Franceschi, R. T., Wilson, J. X. & Dixon, S. J. Requirement for Na⁺-dependent ascorbic acid transport in osteoblast function. *Am. J. Physiol.* **268**, C1430–C1439 (1995).
- Hammarström, L. Autoradiographic studies on the distribution of C¹⁴-labelled ascorbic acid and dehydroascorbic acid. *Acta Physiol. Scand.* **70** (suppl.) 289, 1–75 (1966).
- Rose, R. C. & Bode, A. M. Ocular ascorbate transport and metabolism. *Comp. Biochem. Physiol.* **100**, 273–285 (1991).
- Reiss, G. R., Werness, P. G., Zollman, P. E. & Brubaker, R. F. Ascorbic acid levels in the aqueous humor of nocturnal and diurnal mammals. *Arch. Ophthalmol.* **104**, 753–755 (1986).
- Kodama, T., Kabasawa, I., Tamura, O. & Reddy, V. N. Dynamics of ascorbate in the aqueous humor and tissues surrounding ocular chambers. *Ophthalmic Res.* **17**, 331–337 (1985).
- Romero, M. F., Kanai, Y., Gunshin, H. & Hediger, M. A. Expression cloning using *Xenopus laevis* oocytes. *Methods Enzymol.* **296**, 17–52 (1998).
- Mackenzie, B. in *Biomechanical Transport* (ed. Van Winkle, L. J.) 327–342 (Academic, San Diego, 1999).
- Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank E. Brown for providing MC3T3-E1 cells and P. Fong for providing the pTLN2 vector. This research was supported by the NIH, the National Kidney Foundation, the American Heart Association Massachusetts Affiliate, Cooley's Anemia Foundation, and the International Human Frontier Science Program.

Correspondence and requests for materials should be addressed to M.A.H. (e-mail: mhediger@rics.bwh.harvard.edu). The novel cDNA sequences have been deposited in GenBank with accession numbers AF080452 for rat SVCT1, AF080453 for rat SVCT2, and AF118561 for the rabbit SVCT2 partial sequence.

Regulation of alternative splicing by RNA editing

Susan M. Rueter, T. Renee Dawson & Ronald B. Emeson

Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-6600, USA

The enzyme ADAR2 is a double-stranded RNA-specific adenosine deaminase which is involved in the editing of mammalian messenger RNAs by the site-specific conversion of adenosine to inosine^{1–3}. Here we identify several rat ADAR2 mRNAs produced as a result of two distinct alternative splicing events. One such splicing event uses a proximal 3' acceptor site, adding 47 nucleotides to the ADAR2 coding region, changing the predicted reading frame of the mature ADAR2 transcript. Nucleotide-sequence analysis of ADAR2 genomic DNA revealed the presence of adenosine-adenosine (AA) and adenosine-guanosine (AG) dinucleotides at these proximal and distal alternative 3' acceptor sites, respectively. Use of the proximal 3' acceptor depends upon the ability of ADAR2 to edit its own pre-mRNA, converting the intronic AA to an adenosine-inosine (AI) dinucleotide which effectively mimics the highly conserved AG sequence normally found at 3' splice junctions. Our observations indicate that RNA editing can serve as a mechanism for regulating alternative splicing and they suggest a novel strategy by which ADAR2 can modulate its own expression.

Reverse-transcription polymerase chain reaction (RT-PCR) amplification of total RNA isolated from whole rat brain revealed the expression of four mRNAs encoding double-stranded RNA-specific editase 1 (ADAR2)^{2,4}. Southern-blotting analysis of rat genomic DNA indicated that these mRNAs were derived from a single genomic locus by alternative splicing (data not shown). Sequence analysis of the four isolated rat ADAR2 cDNAs (termed rADAR2a, 2b, 2c and 2f) indicated that these isoforms were produced as a result of two distinct alternative splicing events, one within the region encoding the deaminase domain and the other at the 5' end of the coding sequence (Fig. 1a). Alternative splicing within the deaminase domain results in the exclusion of 30 nucleotides between the second and third putative zinc-coordination motifs (rADAR2a), when compared to the original complementary DNA isolated by Melcher *et al.* (rADAR2b)². This alternative splicing is similar to that previously identified in

human ADAR2 transcripts in which there is a 120-nucleotide insertion at the analogous position^{5,6}. Alternative splicing is also responsible for the production of two additional rat ADAR2 cDNA isoforms containing a 47-nucleotide insertion after the first 28 nucleotides of the coding region (rADAR2e and 2f). Addition of this 47-nucleotide region is predicted to result in a frameshift, producing a protein of relative molecular mass 9K (82 amino acids) that lacks the double-stranded RNA-binding motifs and catalytic deaminase domain required for protein function (Fig. 1a)⁷.

In vitro and tissue-culture model systems have indicated that alternative splicing can modulate the catalytic activity of human ADAR2 protein isoforms without affecting the apparent substrate specificity³. To examine the activity of ADAR2 variants produced as a result of alternative splicing in the rat, we transiently co-transfected human embryonic kidney cells (HEK293) with individual rat ADAR2 cDNAs and a 646-base-pair minigene substrate derived from the B-subunit (GluR-B) of the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) subtype of glutamate receptor. ADAR2 can preferentially modify the Q/R site in GluR-B mRNAs, whereas double-stranded RNA-specific adenosine deaminase (ADAR1) is more selective for the +60 site within intron 11 (refs 2, 8). Primer-extension analysis of RNA transcripts derived from the GluR-B minigene, in the absence of exogenous ADAR expression, revealed minimal editing at either position (Fig. 1b), consistent with previous observations that HEK293 cells contain a low level of endogenous GluR-B editing activity^{2,8,9}. All four rat ADAR2 cDNAs encoded similarly active editing enzymes that selectively modified the Q/R site of GluR-B transcripts and had a significant amount of activity at the +60 site (Fig. 1b). Co-transfection of rat ADAR1 cDNA with the GluR-B reporter construct led to preferential modification at the +60 site, as found previously⁸.

The unexpected observation that rADAR2e and 2f cDNAs encoded functional proteins raised the possibility that ribosomal frameshifting¹⁰ or alternative translation initiation¹¹ could be responsible for the generation of enzymatically active rADAR2 proteins. To distinguish between these possibilities, we transiently transfected expression constructs containing the rADAR2b and 2f cDNAs, with an amino-terminal epitope tag (Flag) inserted directly after the ATG start codon (Met 1), into HEK293 cells. Crude nuclear extracts¹² were analysed by immunoblotting with a polyclonal antiserum raised against the deaminase domain of human ADAR2 (ref. 3) or with a monoclonal antibody directed against the Flag epitope (Fig. 2a). Owing to sequence similarities in the deaminase domain of human (h) ADAR2 (ref. 5) and ADAR1 (refs 13, 14), a 120K immunoreactive species was detected in all samples with the anti-hADAR2 antiserum, which represented endogenous ADAR1 expression; however, no endogenous human ADAR2 protein was detected in non-transfected (control) cells. In HEK293 cells transfected with rADAR2b cDNA, an immunoreactive species migrating at the expected position of full-length epitope-tagged rADAR2 (85K) was detected with both the anti-hADAR2 and anti-Flag antisera, as well as a minor protein species with an apparent relative molecular mass of 80K that was not detected with the anti-Flag antibody. By contrast, cells transfected with the rADAR2f cDNA generated only the 80K species. The identification of an 80K rADAR2 species that was not recognized by anti-Flag antibody indicated that alternative translation initiation or 'leaky scanning'¹⁵ could be responsible for generating the 80K protein isoform, thereby explaining the rADAR2f editing activity evident upon transfection into HEK293 cells (Fig. 1b). Parallel immunoblot analysis of rat-brain nuclear extracts demonstrated that most of the rat ADAR2 immunoreactive material migrated at the expected position for full-length rat ADAR2, with less than 5% of the rat ADAR2 protein corresponding to the truncated isoform (Fig. 2a).

To confirm that rADAR2e and 2f protein expression resulted from internal translation initiation, we used site-directed mutagenesis to mutate each of the two in-frame methionine residues

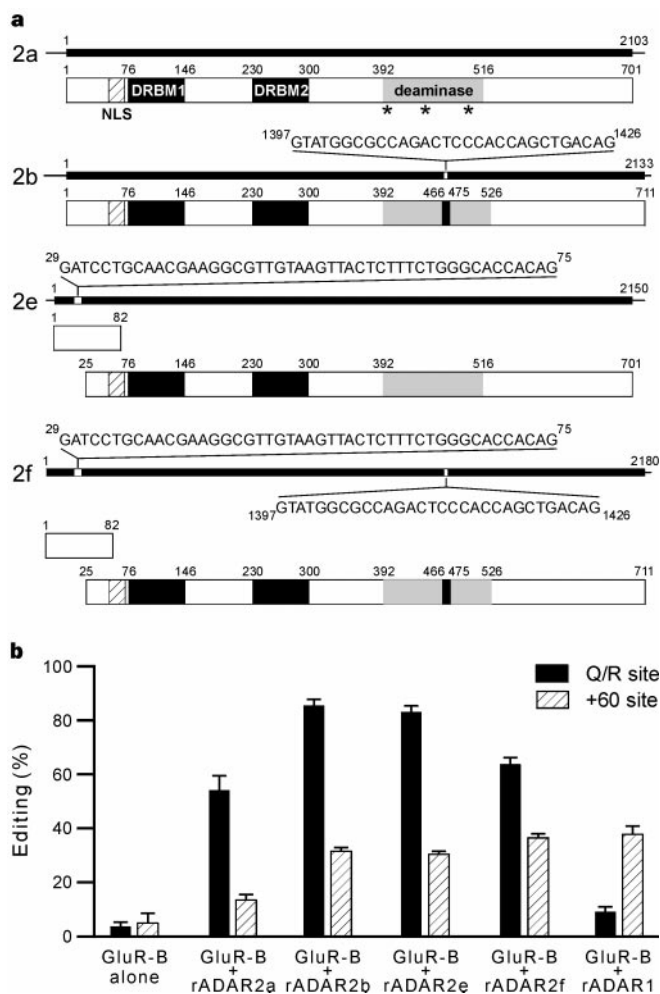


Figure 1 Generation of multiple rADAR2 mRNA and protein isoforms by alternative splicing. **a**, Diagram of four rADAR2 cDNAs (rADAR2a, 2b, 2e and 2f) and their predicted protein products. The positions of the putative bipartite nuclear localization signal (NLS) sequence, double-stranded RNA-binding motifs (DRBM), the catalytic adenosine deaminase domain, putative zinc-coordination motifs (asterisks), and a 10-amino-acid deaminase domain insertion, produced via alternative splicing, are indicated by their coordinates relative to the initiator methionine in rADAR2a. An 82-amino-acid protein is predicted from the rADAR2e and 2f cDNAs, by insertion of 47 nucleotides into the rADAR2 coding region. The sequence and position of the nucleotide insertions within rADAR2 cDNAs are indicated with coordinates relative to the start codon in the rADAR2a cDNA. **b**, Analysis of rADAR2 isoform activity with a GluR-B pre-mRNA substrate. Quantitative analysis of RNA editing of the Q/R (black bars) and +60 (striped bars) sites of a GluR-B RNA substrate is presented from HEK293 cells transiently co-transfected with a GluR-B minigene and individual rADAR2 cDNA isoforms or rADAR1 (mean $\pm$ s.d.; $n = 3$).

(Met 25 or Met 84) most likely to serve as alternative translation start sites. Western-blotting analysis of mutant rADAR2b and 2f proteins with anti-hADAR2 and anti-Flag antisera (Fig. 2b) demonstrated that, although mutation of Met 84 had no effect on the protein expression of either rADAR2 isoform, the M25R mutation prevented rADAR2f expression, leaving rADAR2b synthesis intact. These results indicate that functional rADAR2e and 2f deaminase expression results from internal translation initiation at Met 25, producing proteins lacking the amino-terminal 24 amino acids found in the rADAR2a and 2b isoforms (Fig. 1a).

Rat ADAR2 is expressed in a tissue-specific manner, with mRNA levels being greatest in the brain^{5,16}. To investigate whether alternative splicing of rat ADAR2 mRNAs was also differentially regulated, we used a ribonuclease (RNase) protection strategy to

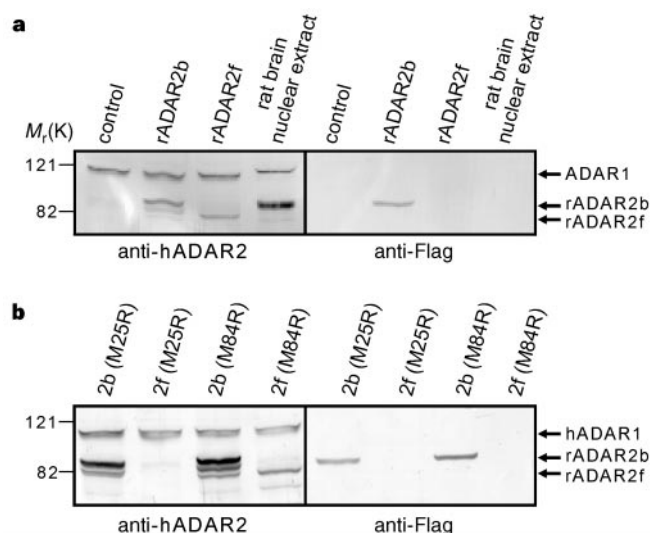


Figure 2 Immunoblotting analysis of epitope (Flag)-tagged wild-type and mutant rADAR2 proteins. **a**, Western-blot analysis of nuclear extracts prepared from non-transfected HEK293 cells (control), HEK293 cells transiently transfected with rADAR2b or rADAR2f cDNAs, or from rat brain using an antiserum directed against the deaminase domain of hADAR2 (ref. 3) or an anti-Flag monoclonal antibody. The anti-hADAR2 antiserum recognizes endogenous ADAR1 protein and rADAR2 in rat brain and transfected HEK293 cells. **b**, Western blot of nuclear extracts prepared from HEK293 cells transiently transfected with rADAR2b and rADAR2f mutants containing an arginine residue replacing the endogenous methionine moieties at positions 25 (M25R) or 84 (M84R). Relative molecular mass markers are indicated to the left of each blot and the identity of ADAR proteins is indicated to the right.

examine alternative splicing patterns in RNAs isolated from various rat tissues. We found that 70–80% of the total rat ADAR2 transcripts in brain and lung contained the 47-nucleotide cassette, whereas only 15–30% were alternatively spliced in this region in the testes and heart (Fig. 3a). Insertion of 30 nucleotides in the deaminase domain also varied between tissues (Fig. 3a). This analysis also revealed an apparent disparity between the expression of ADAR2 mRNA and the protein isoform in rat brain, as most of the ADAR2 transcripts contained the 47-nucleotide cassette (Fig. 3a), whereas immunoblotting indicated that only 5% of the rat ADAR2 immunoreactive material corresponded to the truncated protein isoforms (Fig. 2a). RT-PCR analysis of splicing patterns showed that ~80% of the ADAR2 transcripts from rat brain contained the alternatively spliced cassette (Fig. 3b), as in previous RNase-protection analysis (Fig. 3a). The degree of alternative splicing was comparable for ADAR2 transcripts from mouse brain, whereas inclusion of the 47-nucleotide region in human brain ADAR2 mRNAs was significantly less (11%). Splicing was negligible in any of the tissue-culture cell lines tested, regardless of origin (Fig. 3b).

To explore the mechanisms mediating tissue-specific alternative splicing of rat ADAR2 transcripts further, we isolated genomic fragments containing the alternatively spliced regions by PCR amplification of rat genomic DNA and sequenced them. We found that alternative splicing within the deaminase domain resulted from competition between two 5' donor sites for a common 3' splice acceptor (Fig. 4a, bottom). Both donor sites loosely conformed to the metazoan consensus sequence GTRAGT, in which the initial GT dinucleotide is almost invariant^{17,18}. However, alternative splicing at the 5' end of the coding region resulted from competition between two 3' acceptor sites (Fig. 4a, top). The distal acceptor site, which excluded the 47-nucleotide region, agreed with the conserved consensus sequence (Y)_nXCAG¹⁷, whereas the more proximal 3' splice site included a non-canonical AA dinucleo-

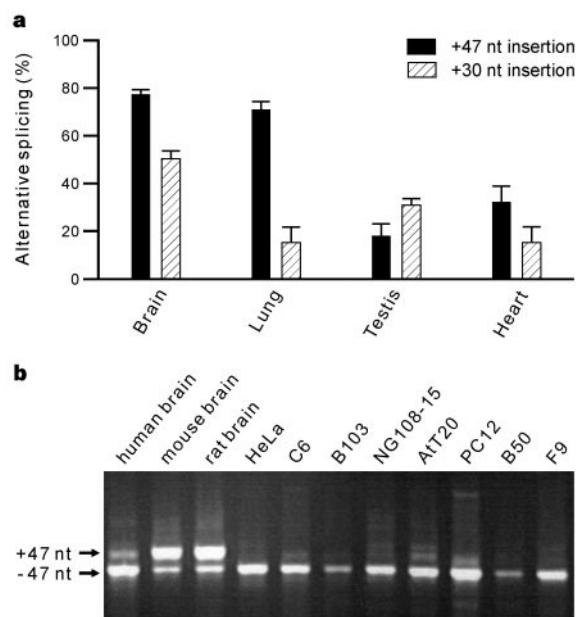


Figure 3 Analysis of tissue-specific ADAR2 alternative splicing patterns. **a**, Summary of results from RNase protection analyses indicating the percentage of total rADAR2 transcripts containing the 47-nt (black bars) or 30-nt (striped bars) insert in different rat tissues (mean $\pm$ s.d.; $n = 3$). **b**, Analysis of ADAR2 alternative splicing in whole mammalian brain samples and tissue-culture cell lines using an RT-PCR-based strategy. The migration positions of PCR reaction products corresponding to transcripts with and without the 47-nt insert are indicated. HeLa, human epithelial; C6, rat glioma; B103, rat neuronal; NG108-15, rat neuroblastoma-glioma; AT20, mouse corticotrope; PC12, rat pheochromocytoma; B50, rat neuronal; F9, mouse teratocarcinoma.

tide at the 3' end of the intron (Fig. 4a, top). Previous studies of splicing *in vitro* have suggested that AA cannot serve as an effective 3' splice site, inhibiting splicing and spliceosome assembly at the first step of transesterification^{19,20}.

As the proximal 3' splice site was preferentially used in rat brain, although it contained a non-functional acceptor site, we considered that RNA editing might selectively modify the rat ADAR2 pre-mRNA sequence to create a competent 3' splice junction. Because of the similar base pairing properties of inosine and guanosine, editing of an adenosine residue at the -1 position (relative to the 47-nucleotide insertion) could result in an AI dinucleotide at the proximal 3' acceptor site which would mimic the highly conserved AG sequence²¹. Sequence comparison of rat ADAR2 genomic DNA and rat-brain cDNA clones, generated by RT-PCR amplification of ADAR2 pre-mRNA, revealed the presence of five adenosine-to-guanosine discrepancies in the region surrounding the 47-nucleotide cassette (Fig. 4b). In addition to three putative editing sites within the insertion itself (at positions +10, +23 and +24), two intronic sites (-2 and -1) were also subject to this apparent post-transcriptional modification (Fig. 4b). These A-to-G disparities were indicative of A-to-I editing events, as observed previously for transcripts encoding the hepatitis delta virus, the AMPA and kainate subtypes of the ionotropic glutamate receptor, and the 2C subtype of the serotonin receptor¹.

To test whether RNA editing could generate a functional 3' splice site, a 3.6-kilobase (kb) ADAR2 rat minigene (as in Fig. 4a, top) was transiently co-transfected into HEK293 cells in the presence or absence of rADAR2b cDNA. RNA editing at the proximal 3' splice site was quantified by primer-extension analysis and splicing patterns were determined by RT-PCR amplification as for Fig. 3b. Cells transfected with the rat ADAR2 minigene alone had background levels of editing in RNA species derived from the transgene, with no detectable alternative splicing (Fig. 4c). Co-transfection of the

rat ADAR2 minigene and rADAR2b cDNA caused a dramatic increase in the editing of transgene-derived RNAs (~30% at the -1 position) and an almost complete inclusion of the alternatively spliced 47-nucleotide region. By comparison, control RNA from whole rat brain also showed that ~30% of ADAR2 pre-mRNAs contained an apparent AG (AI) dinucleotide at the proximal 3' acceptor site and that 80% of the transcripts included the cassette (Fig. 4c). The difference in steady-state levels of editing at the proximal 3' splice junction and inclusion of the 47-nucleotide region in rat mature ADAR2 mRNAs may result from the different kinetics of splicing and editing of the pre-mRNAs. A small proportion of rat ADAR2 pre-mRNAs contained a GG (II) sequence at the proximal 3' splice junction that would not be expected to serve as a functional acceptor site. Analysis of endogenous ADAR2 transcripts in the rat C6 glioma cell line revealed extremely low levels of editing at the proximal 3' acceptor and no inclusion of the 47-nucleotide cassette. These results show that rat ADAR2 can selectively modify its own pre-mRNA to generate a functional 3'-

splice acceptor (AI), thereby regulating the expression of multiple, alternatively spliced ADAR2 transcripts.

An RNA duplex structure is required for A-to-I conversions in the edited region of modified RNA substrates¹. To identify the sequence determinants required for ADAR2 editing, a series of minigenes containing 5' serial deletions (Fig. 5a) plus the rADAR2b cDNA were transiently co-transfected into HEK293 cells and the extent of editing at the proximal 3' splice site was quantified by primer-extension analysis (Fig. 5b). We found that intronic regulatory information critical for editing was contained between -1,668 and -1,154, relative to the proximal 3' splice site. Further analysis of rADAR2 pre-mRNA sequences using RNA-folding algorithms²² identified an imperfect inverted repeat (Fig. 5c) with the potential to form a putative RNA duplex as a result of base-pairing interactions between the upstream region and the 3' end of the intron.

We have identified four rat ADAR2 transcripts that are produced by tissue-specific alternative splicing. One of these splicing events results in the inclusion of an additional 47 nucleotides at the 5' end of the coding region, changing the reading frame of the mature ADAR2 transcript. The inclusion of this 47-nucleotide cassette depends on rat ADAR2 being able to edit its own pre-mRNA in order to generate a new 3' acceptor site by converting an intronic adenosine to an inosine; this inosine can mimic the N1-carbonyl symmetric interaction that occurs between intron terminal residues during pre-mRNA splicing^{20,21}. In rat brain, the high percentage of

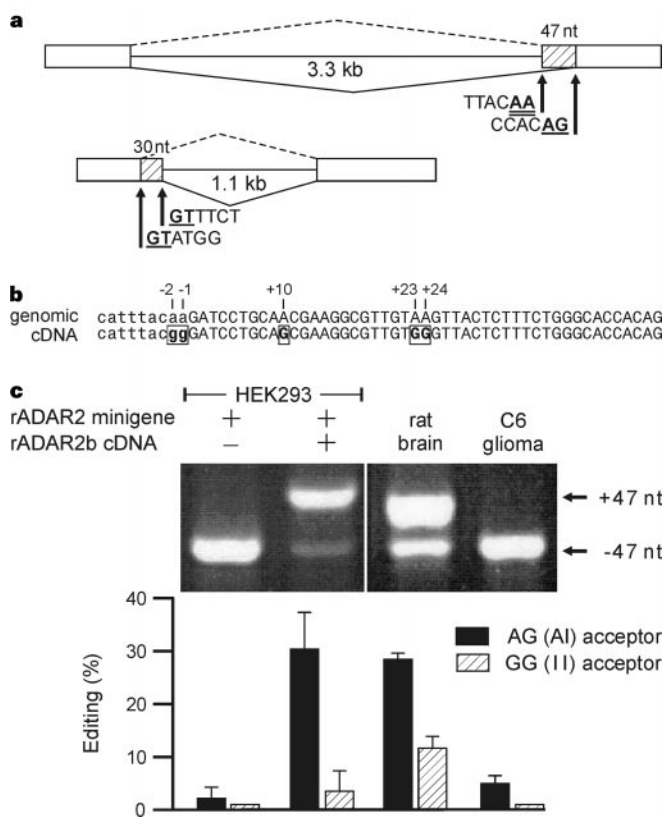


Figure 4 RNA editing and splicing of rat ADAR2 transcripts. **a**, Diagram indicating the structure and alternative splicing patterns of the rADAR2 gene in the regions surrounding the 47-nt (top) and 30-nt (bottom) cassettes. Intronic nucleotide sequences at the alternative 3' and 5'-splice junctions are presented and the highly conserved terminal dinucleotides for the splice donor (GT) and splice acceptor (AG) sites are underlined. The non-canonical AA acceptor site, used for inclusion of the 47-nt cassette, is doubly underlined. **b**, Sequence comparison between rADAR2 genomic DNA and cDNA generated from rADAR2 pre-mRNA. The positions of five putative editing sites (-2, -1, +10, +23 and +24) are indicated by boxed letters; coordinates are relative to the beginning of the 47-nt insertion (upper case letters). **c**, Association of rADAR2 alternative splicing with RNA editing. An analysis of rADAR2 alternative splicing patterns is shown (top) for total RNAs isolated from rat brain, rat C6 glioma cells or HEK293 cells transiently transfected with a rat ADAR2 minigene (like **a**, top) in the presence or absence of rADAR2b cDNA. The migration positions for PCR products with (+47 nt) or without (-47 nt) the alternatively spliced region are indicated. A quantitative analysis of editing is shown below, in which the percentage of total rADAR2 pre-mRNA transcripts containing an AI (black bars) or II (striped bars) dinucleotide at the proximal 3'-splice acceptor site is indicated (mean $\pm$ s.d.; $n = 3$).

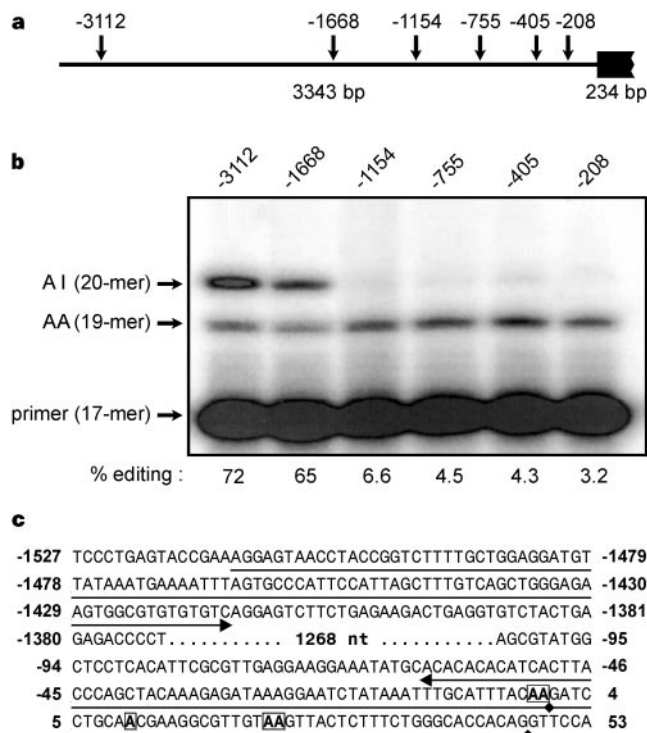


Figure 5 Identification of rADAR2 *cis*-active regulatory elements. **a**, Diagram of rADAR2 minigenes containing 5' serial deletions; intronic coordinates for the 5' end of each construct, total intron length and the length of downstream exon information are indicated. **b**, Primer extension analysis of total RNA from HEK293 cells transiently co-transfected with rADAR2 minigene deletion constructs and rADAR2b cDNA. The migration positions and identity of primer-extension products are indicated as well as a quantitative analysis of editing at the proximal 3' splice site (-1 position). **c**, Nucleotide sequence analysis of rat ADAR2 genomic DNA in the region surrounding the alternatively spliced 47-nt cassette. The positions of five editing sites are indicated by boxed letters and an imperfect inverted repeat is shown with underlying arrows; all coordinates are relative the proximal 3' splice site and the positions of the competing 3' splice junctions are indicated (diamonds).

ADAR2 mRNAs containing the 47-nucleotide insert (Fig. 3a) and the low level of corresponding truncated protein (Fig. 2a) could result from inefficient translation initiation at the downstream Met 25 codon, instability of the truncated ADAR2 proteins, or differential protein localization. Immunocytochemical analysis and immunoblotting of nuclear and cytoplasmic extracts from transfected HEK293 cells indicated that both rADAR2b and 2f are predominantly localized in the nucleus (data not shown). Rapid protein turnover and the low steady-state levels of rADAR2e and 2f (Fig. 2a) could result from the unmasking of a putative degradation signal (PEST)²³ in the truncated proteins (amino acids 20–43). However, restoration of the original rADAR2f reading frame by addition of a single nucleotide generates a full-length 85K protein product (data not shown), suggesting that most ADAR2 translation initiation begins at the Met 1 codon and that downstream translation initiation is inherently inefficient.

Insertion of the 47-nucleotide cassette may be a regulatory mechanism by which rADAR2 protein expression is decreased. Overexpression of rADAR2 in tissue-culture systems can lead to aberrant editing of adenosine residues that are normally not modified *in vivo* (L. K. Hutchinson and R.B.E., unpublished results). The ability of rADAR2 to modulate alternative splicing may represent a negative autoregulatory mechanism by which rADAR2 can prevent its own overexpression in order to avoid editing at aberrant sites. To our knowledge, the splicing of rat ADAR2 is the first example, not only of a naturally occurring inosine at a functional splice junction, but also of alternative splicing regulation by RNA editing. □

Methods

Isolation of rADAR2 cDNA clones. cDNA was generated by RT-PCR amplification of rat-brain total RNA using AMV reverse transcriptase and an oligo(dT) primer. cDNAs were amplified by PCR with sense and antisense oligonucleotides corresponding to the 5'-UTR (–108 to –85) and 3'-UTR (+2,300 to +2,322) of rADAR2; all coordinates of primers are relative to the start codon in rADAR2b (ref. 2), unless otherwise noted. The resultant 2.3-kb PCR reaction product was subcloned into pBluescript SK[–] (Stratagene) and 83 individual cDNA isolates were characterized.

Genomic blotting analysis. Rat genomic DNA was digested with restriction endonucleases, resolved by agarose gel electrophoresis and transferred to a nylon membrane (Zeta-Probe; Bio-Rad) as described²⁴. Probes for Southern blotting were generated by random priming (Prime-It Kit; Stratagene) of a 461-bp rADAR2 cDNA fragment (–2 to +459); hybridization and washing conditions were as described²⁴.

Tissue culture and transfection. HEK293 cells were transiently co-transfected by calcium phosphate precipitation²⁴ with individual rADAR2 cDNA isoforms, either alone or with a 646-nucleotide GluR-B minigene²⁵. For some studies, rADAR2b cDNA was co-transfected with either a 3.6-kb 2-exon rADAR2 minigene (Fig. 4a, top) or 5' serially deleted minigene constructs generated by PCR amplification of rADAR2 genomic DNA (Fig. 5c). All constructs were expressed in the eukaryotic vector pRC-CMV (Invitrogen). Cells were collected ~60 h after transfection and total RNA was prepared by the guanidinium isothiocyanate method and centrifugation through caesium chloride²⁴, followed by treatment with RNase-free DNase (Promega).

Quantification of RNA editing and alternative splicing. For measurements of GluR-B editing, total RNA from transiently transfected HEK293 cells was amplified by RT-PCR using a GluR-B intronic antisense primer (+371 to +390 relative to the Q/R site) and a sense primer specific for the pRC-CMV expression vector. The extent of editing at the Q/R and +60 sites was quantified by primer-extension analysis as described²⁵.

For measurements of rADAR2 editing, total RNA from whole rat brain or transiently transfected HEK293 cells was amplified by RT-PCR using an intronic rADAR2 sense primer upstream from the proximal 3'-acceptor site and an antisense primer specific to either rADAR2 or to the rADAR2 minigene vector to examine endogenous editing or editing in RNAs derived from the rADAR2 transgene in transfected cells, respectively. The PCR reaction product (360 bp for the endogenous rADAR2 gene; 386 bp for the transfected rADAR2 minigene) was purified and editing at the proximal splice acceptor site was

quantified by primer-extension analysis²⁵ with an antisense primer (5'-CGCCTTCGTTGCAGGAT-3') extended in the presence of 1.2 mM dATP, 1.2 mM dCTP, 1.2 mM dGTP and 5 mM dideoxythymidine 5'-triphosphate followed by phosphorimager analysis (Molecular Dynamics).

Quantification of rat ADAR2 alternative splicing products was performed after RT-PCR amplification of total RNA isolated from whole brain, tissue-culture cell lines or transfected HEK293 cells, by using an ADAR2 sense primer in the 5' untranslated region and an antisense primer specific to either ADAR2 (for endogenous RNA splicing) or to the transfected ADAR2 minigene vector downstream from the 47-nt insertion. The percentage of rat ADAR2 transcripts containing the 47-nt region was quantified by phosphorimager analysis (Molecular Dynamics).

Western-blotting analysis. Crude nuclear extracts were prepared from HEK293 cells¹² and immunoblotted as described²⁴ with either an anti-hADAR2 polyclonal antiserum directed against the deaminase domain³ (gift from M. O'Connell and W. Keller) or with an anti-Flag M2 monoclonal antibody (Kodak), followed by secondary antisera conjugated to alkaline phosphatase (Bio-Rad). For colorimetric detection, 5-bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium were used as substrates for alkaline phosphatase²⁴.

Ribonuclease protection analysis. Total RNA was prepared from various rat tissues and alternative splicing was analysed by RNase protection as described²⁶. For construction of the RNase-protection probe, a 342-bp region from the rADAR2e cDNA (–34 to 308) was fused to a 397-bp fragment from the rADAR2a deaminase domain (+1,125 to +1,521). A 130-nt rat cyclophilin probe was included with the rADAR2 probe to normalize for the amount of RNA in each hybridization reaction. Protected RNA fragments were separated on a 4% acrylamide/7M urea gel and quantified by phosphorimager analysis.

Site-directed mutagenesis. Two in-frame methionine codons encoding Met 25 and Met 84 were mutated to arginine codons (AGA) by oligonucleotide-directed site mutagenesis in pBSK[–] (Stratagene) as described²⁷ using synthetic oligonucleotides to introduce *AccI* restriction sites at nucleotide positions 67–72 and 244–249 relative to the normal translation start site of rADAR2b (ref. 2).

Received 11 November 1998; accepted 11 March 1999.

- Rueter, S. & Emeson, R. in *Modification and Editing of RNA* (eds Grosjean, H. & Benne, R.) 343–361 (ASM, Washington DC, 1998).
- Melcher, T. *et al.* A mammalian RNA editing enzyme. *Nature* **379**, 460–464 (1996).
- O'Connell, M. A., Gerber, A. & Keller, W. Purification of human double-stranded RNA-specific editase 1 (hRED1) involved in editing of brain glutamate receptor B pre-mRNA. *J. Biol. Chem.* **272**, 473–478 (1997).
- Bass, B. L. *et al.* A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* **3**, 947–949 (1997).
- Gerber, A., O'Connell, M. A. & Keller, W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* **3**, 453–463 (1997).
- Lai, F., Chen, C. X., Carter, K. C. & Nishikura, K. Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol. Cell. Biol.* **17**, 2413–2424 (1997).
- Lai, F., Drakas, R. & Nishikura, K. Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. *J. Biol. Chem.* **270**, 17096–17105 (1995).
- Maas, S. *et al.* Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.* **271**, 12221–12226 (1996).
- Burns, C. M. *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303–308 (1997).
- Farabaugh, P. J. Programmed translational frameshifting. *Microbiol. Rev.* **60**, 103–134 (1996).
- Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I. & Santonno, C. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proc. Natl Acad. Sci. USA* **92**, 11598–11602 (1995).
- Schreiber, E., Matthias, P., Muller, M. M. & Schaffner, W. Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. *Nucleic Acids Res.* **17**, 6419 (1989).
- Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc. Natl Acad. Sci. USA* **91**, 11457–11461 (1994).
- O'Connell, M. A. *et al.* Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol. Cell Biol.* **15**, 1389–1397 (1995).
- Kozak, M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887–903 (1991).
- Melcher, T. *et al.* RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J. Biol. Chem.* **271**, 31795–31798 (1996).
- Mount, S. M. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Jackson, I. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**, 3795–3798 (1991).
- Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
- Deirdre, A., Scadden, J. & Smith, C. W. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14**, 3236–3246 (1995).
- Tarn, W. Y. Site-specific substitution of inosine at the terminal positions of a pre-mRNA intron: implications for the configuration of the terminal base interaction. *Biochimie* **78**, 1057–1065 (1996).
- Zuker, M. Prediction of RNA secondary structure by energy minimization. *Methods Mol. Biol.* **25**, 267–294 (1994).
- Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267–271 (1996).

24. *Current Protocols in Molecular Biology* (eds Ausubel, F. et al.) (Wiley, New York, 1989).
 25. Rueter, S. M., Burns, C. M., Coode, S. A., Mookherjee, P. & Emeson, R. B. Glutamate receptor RNA editing *in vitro* by enzymatic conversion of adenosine to inosine. *Science* **267**, 1491–1494 (1995).
 26. Emeson, R. B., Hedjran, F., Yeakley, J. M., Guise, J. W. & Rosenfeld, M. G. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. *Nature* **341**, 76–80 (1989).
 27. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488–492 (1985).

Acknowledgements. We thank J. Barnett, J. Patton, B. Wadzinski, C. Desai and members of R.B.E.'s laboratory for critically reading this manuscript. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to R.B.E. (e-mail: ron.emeson@mcmill.vanderbilt.edu).

Structural basis for initiation of transcription from an RNA polymerase–promoter complex

Graham M. T. Cheetham^{†‡}, David Jeruzalmi^{†‡} & Thomas A Steitz^{†§}

Departments of ^{*}Molecular Biophysics and Biochemistry and [§]Chemistry and the [†]Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA

Although the single-polypeptide-chain RNA polymerase from bacteriophage T7 (T7RNAP), like other RNA polymerases, uses the same mechanism of polymerization as the DNA polymerases, it can also recognize a specific promoter sequence, initiate new RNA chains from a single nucleotide, abortively cycle the synthesis of short transcripts, be regulated by a transcription inhibitor, and terminate transcription^{1–3}. As T7RNAP is homologous to the Pol I family of DNA polymerases⁴, the differences between the structure of T7RNAP complexed to substrates and that of the corresponding DNA polymerase complex provides a structural basis for understanding many of these functional differences. T7RNAP initiates RNA synthesis at promoter sequences that are conserved from positions –17 to +6 relative to the start site of transcription. The crystal structure at 2.4 Å resolution of T7RNAP complexed with a 17-base-pair promoter shows that the four base pairs closest to the catalytic active site have melted to form a transcription bubble. The T7 promoter sequence is recognized by interactions in the major groove between an antiparallel β-loop and bases. The amino-terminal domain is involved in promoter recognition and DNA melting. We have also used homology modelling of the priming and incoming nucleoside triphosphates from the T7 DNA-polymerase ternary complex structure to explain the specificity of T7RNAP for ribonucleotides, its ability to initiate from a single nucleotide, and the abortive cycling at the initiation of transcription.

T7RNAP has a relative molecular mass of 99,000 ($M_r = 99K$) and is homologous to the RNA polymerases from the T3 and SP6 phages^{5,6}, as well as the nuclear-encoded mitochondrial RNA polymerases from fungi, plants and animals⁷. Hydroxyl radical footprinting of the initiation phase of RNA synthesis reveals an extension of DNA protection by T7RNAP downstream from the promoter with no change in upstream protection⁸, a phenomenon explained either by protein expansion ('inchworming') or DNA contraction ('scrunching'). In this initiation phase, the RNA polymerase can abort synthesis of the transcript before entering the elongation phase, resulting in short transcripts (abortive cycling)². The relative simplicity of T7RNAP makes it an attractive polymerase to study^{9,10}.

We co-crystallized T7RNAP with a blunt-ended, 17-base-pair

(bp) promoter DNA (sequence given in Fig. 1). The structure of the complex was determined at 2.4 Å resolution by multiple isomorphous replacement, using selenomethionyl-substituted protein and iodo-containing oligonucleotides, and a mercury acetate derivative. The coordinates of the complex have been refined to yield an R_{free} value of 0.27 (Table 1).

T7RNAP and other members of the DNA-polymerase I family have modular structures, with auxiliary domains and motifs specific to each polymerase attached to the catalytic core domain¹¹. The 324-residue N-terminal domain of T7RNAP, which is unique to the RNA-polymerase members of the Pol-I family, blocks part of the active-site cleft⁹ which binds primer-template product in DNA polymerases. Thirteen base pairs (positions –17 to –5) of the T7 promoter are bound to this N-terminal domain in a position away from the location of primer-template DNA seen in DNA polymerases^{12,13}. Base pairs –1 to –4 are melted by T7RNAP and the template strand is directed into the catalytic active-site cleft of the polymerase.

Sequence-specific recognition of the duplex promoter by T7RNAP is accomplished by a combination of direct and indirect readout in adjacent major and minor grooves, respectively (Fig. 2a). Direct base-recognition is achieved by an anti-parallel β-ribbon that is inserted into the DNA major groove. Most DNA–protein complexes perform sequence-specific recognition of bases by inserting an α-helix, present in a variety of motifs, into the major groove of B-form DNA¹⁴; however, sequence recognition by antiparallel β-strands is far less common^{15,17}.

Although, in principle, up to six amino-acid positions can be presented by antiparallel β-strands for interaction with the exposed edges of base pairs¹⁸, T7RNAP uses four side-chains lying on one side of an extended antiparallel β-hairpin motif, or specificity loop (residues 739–770). These form the basis for discrimination between T7 promoter and other DNA sequences. In contrast to other Pol-I polymerases, this specificity loop is an insertion between the polymerase fingers and palm subdomains and exhibits no known homology to other sequence-dependent DNA-binding motifs.

The T7, T3 and SP6 RNA polymerases are a homologous family of enzymes that recognize and discriminate among similarly homologous but different promoter sequences. This is accomplished mainly by the four protein side chains that interact with four

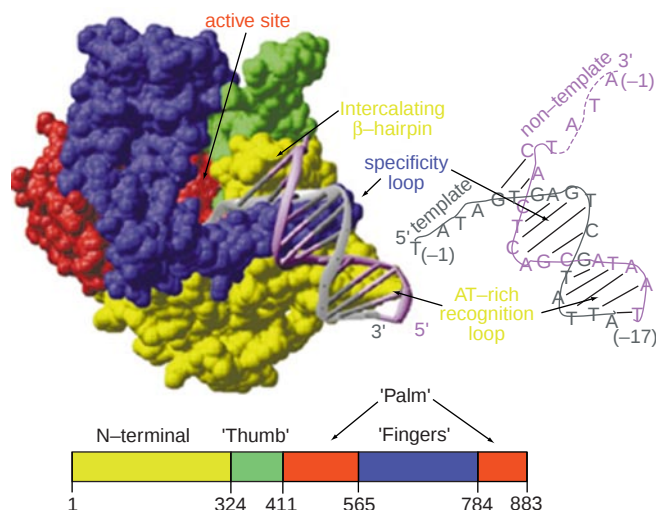


Figure 1 Surface representation of the T7RNAP–promoter complex structure. In all figures, the polymerase subdomains are coloured red 'palm', green 'thumb' and blue 'fingers'. The N-terminal domain which is specific to all phage-like RNA polymerases is yellow, and the template and non-template strands are shown in grey and magenta, respectively. The sequence of the promoter and sequence organization of the T7RNAP domains are also given.

[‡] Present address: Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

bases. The guanidinium group of Arg 756 interacts with the 6-keto and 7-imino groups of G-9, while Gln 758 and Arg 746 make bidentate hydrogen-bonds with A-8 and G-7, respectively (Fig. 2b). The only major groove interactions with bases of the non-template strand are made by Asn 748, which directly hydrogen-bonds to N7 and indirectly interacts through a water-mediated hydrogen bond to the 6-keto group of G-11. These interactions are essential for recognition of the T7 promoter; mutation of Asn 748 to Asp makes T7RNAP specific for T3 promoters⁵, explaining why the T3 enzyme has Asp at this position. Similarly, the primary discrimination elements of the SP6 promoter are base pairs at -8 and -9 (ref. 6). The Asp 748 mutant of T7RNAP can recognize the T3 promoter by forming an alternative hydrogen-bonding network with C-11 (non-template) and G-10 (template), and by relocating the water molecule within the protein-DNA interface. The bidentate interaction of Arg 756 and Gln 758 with the template strand at positions A-8 and G-9 accounts for the inability of T7RNAP specifically to recognize and use SP6 promoters, which have thymine and cytosine at these positions. The recognition residues in the SP6 enzyme corresponding to 756 and 758 are lysine and

serine.

The N-terminal domain of T7RNAP indirectly recognizes the DNA sequence in the minor groove at base pairs -17 to -13 by inserting a flexible surface loop, or AT-rich recognition loop (residues 93-101), into the minor groove (Fig. 2a). At the same time, the phosphodiester backbone is distorted to form a wider (7 Å versus 5.7 Å) and shallower (4.4 Å versus 7.5 Å) minor groove, compared to B-form DNA. Thus, T7RNAP may recognize the A + T-rich sequence through its inherent flexibility¹⁴. These distortions cause a slight bend in the axis of the DNA helix. However, compared to the DNA distortion produced by eukaryotic TATA-binding proteins^{19,20}, unwinding of the double helix at these positions by T7RNAP is minimal.

The N-terminal domain also facilitates melting of the promoter duplex, both by inserting a β -hairpin containing Val 237 in place of base pair -4 to stack on base pair -5, and by providing an extensive binding site for the melted template strand (Fig. 2c). The packing of this β -hairpin containing Val 237 is also presumably essential for the maintenance of the upstream part of the bubble during the elongation phase of transcription. An analogous denaturation of duplex

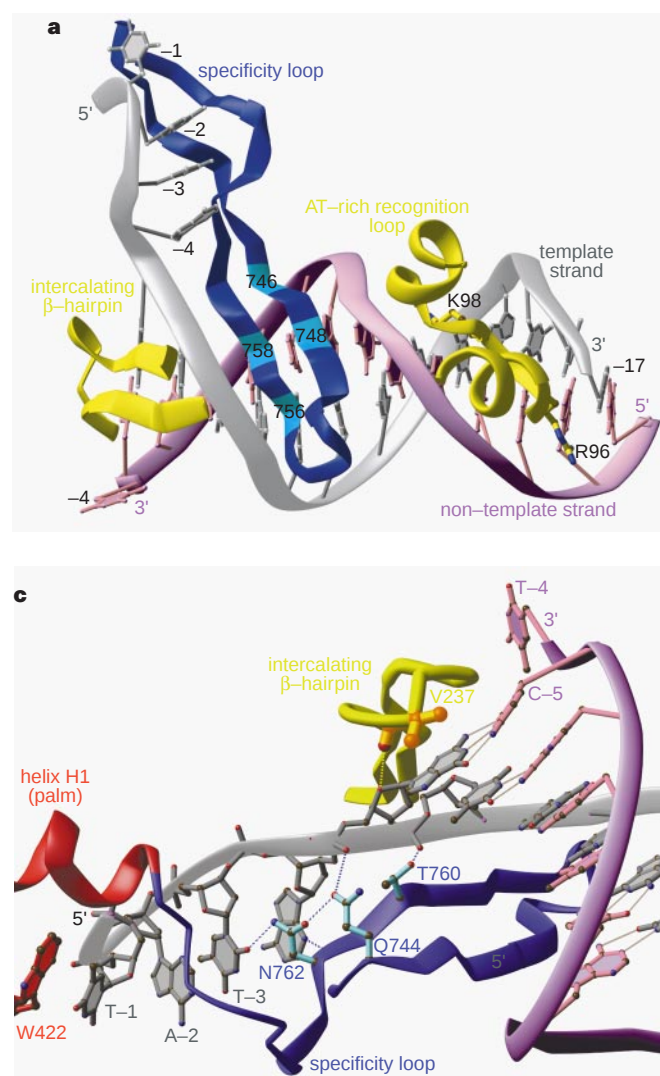
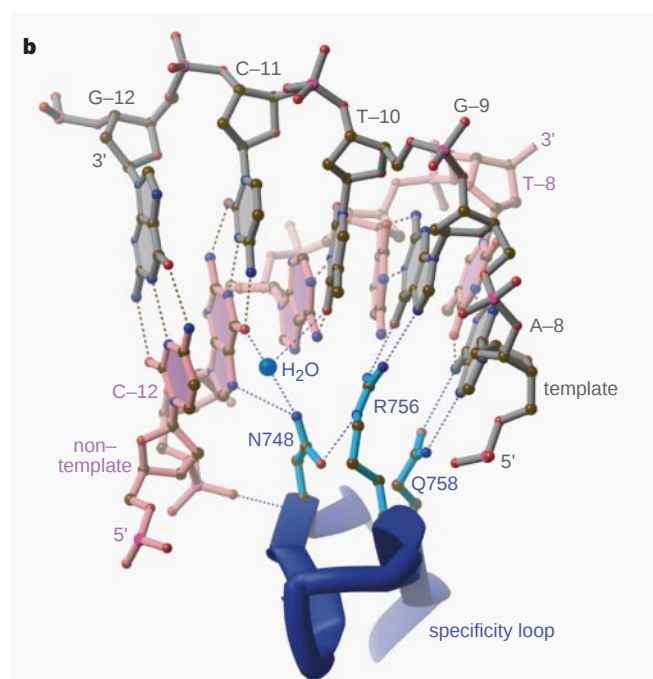


Figure 2 Interactions between the T7-promoter DNA and three regions of T7RNAP that are unique to the RNA polymerase. **a**, Two loops from the N-terminal domain and one from the fingers interact with the promoter. The side chains of Arg 96 and Lys 98 located in the N-terminal AT-rich minor-groove-recognition loop are shown in detail, and the C α positions of the four side chains that interact in the major groove are indicated. **b**, Sequence-specific hydrogen-bonding interactions between protein side chains and bases in the major groove of the promoter. A



bound water molecule is part of the polymerase-DNA interface, as are the three side chains shown. **c**, Formation of the upstream transcription bubble facilitated by duplex disruption and template binding. Residue Val 237 separates the template and non-template strands and redirects the exposed template towards the catalytic active site. The 5' terminus of the template strand (position -1) is anchored in a pocket at the junctions of the palm, fingers and N-terminal domains by a base-stacking interaction with residue Trp 422.

RNA occurs in the complex between tRNA^{Gln} and Gln-tRNA synthetase, where Leu 136 stacks on G2-C71 and melts base pair U1-A72 (ref. 21).

The promoter is further stabilized in an open conformation by interactions between the single-stranded portion of the template strand and residues belonging to the specificity loop (Fig. 2c). These interactions are mostly confined to the phosphodiester backbone, consistent with the need for sequence-independent binding of the template strand during translocation. The aromatic side chain of Trp 422 stacks on the -1 base, inducing a sharp bend in the template strand, which orients the base at position +1 for initiating ribonucleotides.

Because the T7 RNA and DNA polymerases are homologous enzymes of the Pol-I family, the structure of the DNA polymerase complexed with primer-template and dGTP¹³ can be used to model by homology the +1 to +3 template strand as well as the priming and incoming ribonucleoside triphosphates onto the T7RNAP. To align the two polymerases, 17 α -carbon atoms from the three β -strands in the polymerase palm domains containing the conserved carboxylates in each enzyme were superimposed with a root-mean-square deviation (r.m.s.d.) of 0.67 Å. We assumed that the substrates bind to both polymerases identically, and transferred five nucleotides at the primer terminus of the DNA-polymerase complex to the T7RNAP (Fig. 3). This produced a model for the +1 to +3 template nucleotides, the priming GTP and the incoming GTP. The β - and α -phosphates of the priming GTP were positioned to make plausible interactions with the T7RNAP protein.

The homology-modelled initiation complex fits well onto the T7RNAP promoter complex, with no steric clashes; it explains many of the functional properties of T7RNAP. The surface topology of the RNA polymerase exactly accommodates the +1 to +3 nucleotides of the template strand in the conformation observed in the DNA polymerase (Fig. 3), supporting the idea that these two enzymes bind this portion of the template in the same way. The two Mg²⁺ ions and the incoming nucleoside triphosphate are positioned next to the catalytic carboxylates in exactly the same way in both enzymes. Perhaps the strongest support for this model is the distance of 1.4 Å between the terminal 5' phosphate (position -1) in the RNA polymerase co-crystal structure and the O3' ribose position (+1) homology modelled from the T7DNAP complex, without any optimization of their relative positions.

The model suggests two reasons why T7RNAP can initiate

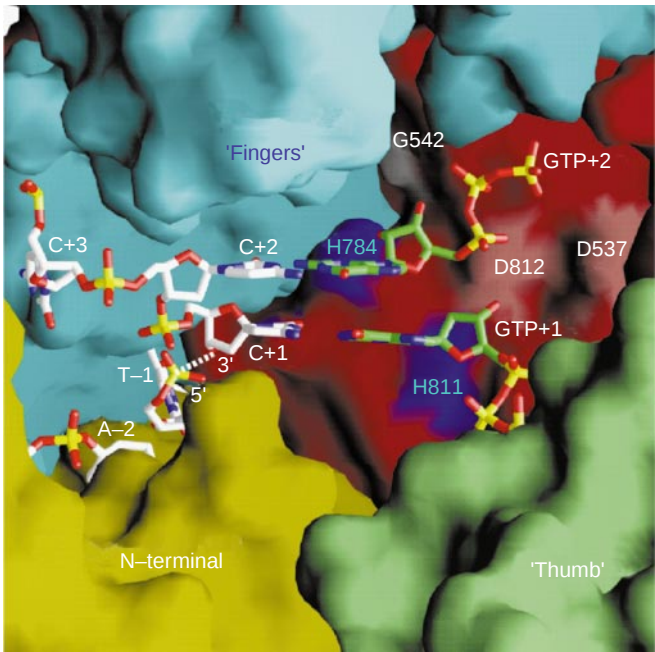


Figure 3 Homology-modelled initiation complex. A surface representation of the active-site cleft of T7RNAP. Nucleotides -2 and -1 of the T7RNAP template strand observed in this complex, the homology-model-built priming (+1) and incoming (+2) GTPs and the complementary template strand (+1 to +3) derived from the T7 DNA-polymerase ternary complex structure¹³ are shown in detail. The areas of histidine residues 784 and 811 are coloured cyan, residue Gly 542 is grey, and residues Asp 812 and Asp 537 are shown in pink.

synthesis from a single priming nucleotide, whereas the DNA polymerase cannot. First, as a result of the specific promoter binding, the RNA polymerase exactly positions the template strand in the active site, providing important complementary base-pairing sites for the bases of the priming and incoming nucleoside triphosphates at the +1 and +2 positions, unlike a single-stranded DNA bound at DNA-polymerase active sites. Second, there is an additional interaction between the 2'-hydroxyl of the priming nucleotide and the RNA polymerase that would further stabilize the positioning of this nucleotide. The N δ 1 posi-

Table 1 Summary of crystallographic analysis

Data collection and MIRAS analysis															
	Native		Hgl		Iodine I		Iodine II		SeMet						
Wavelength (Å)	0.92		1.5418		1.5418		0.92		0.979						
Resolution (Å)	2.4		3.4		3.8		4.0		3.2						
Completeness	96.7		91.3		85.3		91.4		79.0						
R sym (I)*	0.08		0.12		0.12		0.07		0.08						
No. of sites			5Hg + 2I		1		2		22						
R _{iso} (F)†			0.32		0.17		0.25		0.18						
Phasing power			1.21		0.67		0.84		1.03						
acentric data‡															
Solvent-flattening density modification															
Resolution	20–15.0	10.0	7.0	5.5	4.8	4.4	4.0	3.8	3.6	3.4	3.2	3.0	2.8	All	
MIRAS F.o.m.§ all data	0.83	0.7	0.7	0.7	0.61	0.5	0.4	0.4						0.52	
SOLO F.o.m. all data	0.86	0.9	0.8	0.8	0.8	0.8	0.8	0.83	0.8	0.71	0.6	0.5	0.4	0.5	
Structure refinement (40–2.4 Å)															
Resolution (Å)	40–2.4 <i>F</i> > 3σ(<i>F</i>)														
No. reflections working/free	45,066/5,011														
<i>R</i> / <i>R</i> _{free} 40–2.40 Å (outer bin 2.45–2.40 Å)	0.22 (0.32)/0.27 (0.37)														
r.m.s. d. (bond), (angle)¶	0.009 Å, 1.47°														
No. protein residues/nucleotides/water	862 (missing residues 1–5, 56–71)/31 (missing non-template –1 to –3)/450														

* $R_{\text{sym}} = \sum (I - \langle I \rangle) / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity for multiple observations of symmetry related reflections.
† Mean fractional isomorphous difference = $\sum |F_p| |F_c| / \sum |F_p|$ where $|F_p|$ is the protein structure factor amplitude and $|F_c|$ is the heavy-atom structure factor amplitude.
‡ Phasing power = $r.m.s. (|F_p|/E)$, where E is the residual lack of closure.
§ F.o.m., figure of merit.
|| R_{free} is the same as R_{cryst} but calculated on the 10% of data excluded from refinement. $R = \sum (|F_p| - |F_c|) / \sum (|F_p|)$, where F_p and F_c are the observed and calculated structure factor amplitudes, respectively.
¶ Root-mean-squared deviation given from ideal values.

tion of His 811, which is implicated in ribonucleotide binding and phosphodiester-bond formation²², is ideally positioned to hydrogen-bond with the 2'-OH of the priming GTP. This interaction also assures initiation by ribo rather than deoxyribonucleotides.

T7RNAP can incorporate ribonucleotides in preference to deoxyribonucleotides because of His 784 and Gly 542. The discrimination of DNA polymerases against ribonucleotides has been proposed²³ to result from these enzymes presenting a 'steric gate' to the incoming nucleotide; the steric gate is a bulky residue that overlaps the position that would be occupied by the 2'-OH of a ribonucleotide. In most DNA polymerases of the Pol-I family, this residue is a glutamic acid (residue 710 in the Klenow fragment). The corresponding homologous residue in T7RNAP is Gly 542, allowing room for a 2'-OH. T7RNAP discriminates against a deoxy substrate partly by providing His 784 to hydrogen-bond to the 2'-OH; desolvating His 784 in the absence of the 2'-OH, leaving a hole, would be energetically costly. Both Gly 542 and His 784 are absolutely conserved in all phage-like DNA-dependent RNA polymerases.

This co-crystal structure indicates that the abortive cycling that occurs at the initiation of transcription requires accumulation of the DNA template strand at the polymerase active site. The extent of promoter protection from hydroxyl-radical hydrolysis produced by methidiumpropyl-EDTA-Fe(II) in the presence of T7RNA polymerase at various early stages in the elongation reaction has been determined⁸. Although protection of the downstream DNA sequences extended as the open complex was formed and RNA was synthesized, the extent of protection of the upstream sequences did not change. At least two models can explain how the polymerase can expand its downstream protection of the DNA while maintaining the upstream protection—'protein inchworming' and 'DNA scrunching'. The protein-inchworming model suggests that the duplex recognition portion of the enzyme and the polymerase portion of the enzyme reside on different domains and can move independently of each other as RNA synthesis proceeds. The DNA-scrunching model keeps the enzyme structure constant but requires that the template and non-template DNA strands can be accommodated progressively in the polymerase active site as synthesis proceeds; the single-stranded DNA is compacted or 'bunched' near the synthesis site. The synthesis active site and the promoter-binding region cannot move relative to each other because the antiparallel β -ribbon that is recognizing the promoter DNA comes from the 'fingers' domain, which also provides part of the binding site for the incoming nucleoside triphosphate. We propose that, during the abortive initiation phase of RNA synthesis, the template strand accumulates in the active site while the promoter duplex is still bound to the protein. □

Methods

We purified T7RNAP as described⁹ from *Escherichia coli* strain BL21(DE3) containing plasmid pAR1219 (ref. 24). The selenomethionyl enzyme was produced in *E. coli* strain B834(DE3) and purified under strict reducing conditions. For crystallization, duplex T7 DNA promoters, made by annealing trityl-purified (Vydac C₄ reversed-phase column) synthetic oligonucleotides, were mixed with T7RNAP in a 1.3:1 ratio at a concentration of 30 mg ml⁻¹ and pH 7.8. We obtained very thin plate-like crystals, which diffract to 2.4 Å resolution at the CHESS F1 beamline, by vapour diffusion of the complex mixed with an equal volume of the well solution against a well solution containing 20% PEG 8000K, 200 mM ammonium sulphate, 0.25% β -octyl glucopyranoside and 100 mM Tris-HCl, pH 8.7. Derivatives were prepared either by adding heavy-atom reagents to the crystallization drop or by using modified DNA oligonucleotides in crystallization, and diffraction data were collected at 100K–110K using crystals that had been stabilized by addition of 16% propylene glycol before flash-freezing in liquid propane. We reduced the intensity data using MOSFLM²⁵, DENZO²⁶ and the CCP4 program suite²⁷ to reveal a unit cell of dimensions $a = 220.1$, $b = 73.3$, $c = 161.9$ Å and space group $P2_12_12_1$. However, subsequent analysis of these data revealed a systematic

distribution of weak intensities for reciprocal lattice with planes odd values of l , resulting from the two molecules in the crystallographic asymmetric unit being separated by a translation corresponding to almost exactly $c/2$. As the mean $I/\sigma(I) = 1.3$, for all odd values of l data, further analyses were performed in a unit cell with one half the c dimension and one molecule per asymmetric unit. Reducing the number of parameters by 2 proved more useful than including the extremely weak reflections. Initial experimental phases calculated by multiple isomorphous replacement with anomalous scattering with the program MLPHARE²⁷ gave readily interpretable electron-density maps only after phase extension from 3.6 Å to 2.8 Å resolution using solvent-flattening density-modification (Program Solomon²⁸). Iterative cycles of model rebuilding²⁹ and refinement³⁰ were performed, initially incorporating experimental phase information in electron-density-map calculation and model-refinement restraints. Individual B -factors were refined and a bulk solvent correction was applied. Coordinates have been deposited in the Protein Data Bank (accession number 1CIZ).

Received 23 December 1998; accepted 9 March 1999.

- Steitz, T. A., Smerdon, S. J., Jaeger, J. & Joyce, C. M. A unified polymerase mechanism for nonhomologous DNA and RNA polymerase. *Science* **266**, 2022–2025 (1994).
- Carpousis, A. J. & Gralla, J. D. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation *in vitro* at the lac UV5 promoter. *Biochemistry* **19**, 3245–3253 (1980).
- McAllister, W. T. & Wu, H.-L. Regulation of transcription of the late genes of bacteriophage T7. *Proc. Natl Acad. Sci. USA* **75**, 804–808 (1978).
- Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461–467 (1990).
- Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228**, 506–515 (1992).
- Lee, S. S. & Kang, C. A two base-pair substitution in T7 promoters by SP6 promoter-specific base-pairs alone abolishes T7 promoter activity but reveals SP6 promoter activity. *Biochem. Int.* **26**, 1–5 (1992).
- Masters, B. S., Stohl, L. L. & Clayton, D. A. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophage T3 and T7. *Cell* **51**, 89–99 (1987).
- Ikeda, R. A. & Richardson, C. C. Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl Acad. Sci. USA* **83**, 3614–3618 (1986).
- Jeruzalmi, D. & Steitz, T. A. Structure of the T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* **17**, 4101–4113 (1998).
- Souza, R., Chung, Y. J., Rose, J. P. & Wang, B.-C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **364**, 595–599 (1993).
- Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).
- Eom, S. H., Wang, J. & Steitz, T. A. Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* **382**, 278–281 (1996).
- Doublé, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
- Steitz, T. A. Structural studies of protein–nucleic acid interactions: The sources of sequence specific binding. *Q. Rev. Biophys.* **23**, 205–280 (1990).
- Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF–DNA complex: A protein induced DNA U-turn. *Cell* **87**, 1295–1306 (1996).
- Somers, W. S. & Phillips, S. E. V. Crystal structure of the met repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β -strands. *Nature* **359**, 387–393 (1992).
- Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. DNA recognition by beta-sheets in the Arc repressor–operator crystal structure. *Nature* **367**, 754–757 (1994).
- Suzuki, M. DNA recognition by a β -sheet. *Protein Eng.* **8**, 1–4 (1995).
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
- Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
- Rould, M. A. & Steitz, T. A. Structure of the glutamyl-tRNA synthetase-tRNA^{Glu}-ATP complex. *Nucleic Acids Mol. Biol.* **6**, 225–245 (1992).
- Osumi-Davis, P. A., Aguilera, M. C., Woody, R. W. & Woody, A. Y. M. Asp 357, Asp 812 are essential and Lys 631, His 811 are catalytically significant in bacteriophage T7 RNA polymerase activity. *J. Mol. Biol.* **226**, 37–45 (1992).
- Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A. & Goff, S. P. Conferring RNA polymerase activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA* **94**, 407–411 (1997).
- Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **81**, 2035–2039 (1984).
- Leslie, A. G. W. Joint CCP4 and ESRF-EACMB Newsletter Protein crystallogr. No. 26 (Daresbury Laboratory, Warrington, UK, 1992).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project No. 4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. et al. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Acknowledgements. We thank many members of the T.A.S. laboratory for assistance with data collection at beamlines X12C (NSLS, Brookhaven Laboratory), A1, F1 and F2 (Cornell High Energy Synchrotron). This work was supported by a grant from the N.I.H.

Correspondence and requests for materials should be addressed to T.A.S.

corrections

Experimental verification of the quasi-unit-cell model of quasicrystal structure

Paul J. Steinhart, H.-C. Jeong, K. Saitoh, M. Tanaka, E. Abe & A. P. Tsai

Nature 396, 55–57 (1998)

Figure 3 of this Letter wrongly shows all cobalt sites as yellow filled circles, indicating that they lie on the same $c = 0$ level, whereas the two lower sites in the figure were intended to be open circles, indicating cobalt sites on the $c = 1/2$ level. A corrected version of the figure can be found on *Nature's* website.

Extreme Th1 bias of invariant V α 24J α Q T cells in type 1 diabetes

S. Brian Wilson, Sally C. Kent, Kurt T. Patton, Tihamer Orban, Richard A. Jackson, Mark Exley, Steven Porcelli, Desmond A. Schatz, Mark A. Atkinson, Steven P. Balk, Jack L. Strominger & David A. Hafler

Nature 391, 177–181 (1998)

In this Letter, we reported that invariant V α 24J α Q T cells from monozygotic diabetic twins/triplets were reduced in number and produced only interferon- γ on appropriate stimulation, whereas those cloned from at-risk non-diabetic twins/triplets and controls produced both interferon- γ and interleukin(IL)-4. We also reported (see our Fig. 4) that 50% (7/14) of high-risk diabetes non-progressors had markedly raised levels of serum IL-4, as measured by ELISA. However, we now find that measurement of serum IL-4 by ELISA is confounded by the presence in some serum samples of a heterophile-like substance(s) that gives false positive estimations for IL-4 by crosslinking the capture and detection antibodies used in the assay^{1,2}. This is not the case when IL-4 is measured in the tissue-culture medium of *in vitro* activated T cells. Details will be published once the source of error is determined, a method for accurate measurement of serum IL-4 is established, and the apparent association of production of the heterophile substance and/or IL-4 with diabetic non-progression has been clarified. □

1. Redondo, M. *et al.* *Diabetes* 48 (suppl. 1) Abstr. (1999).

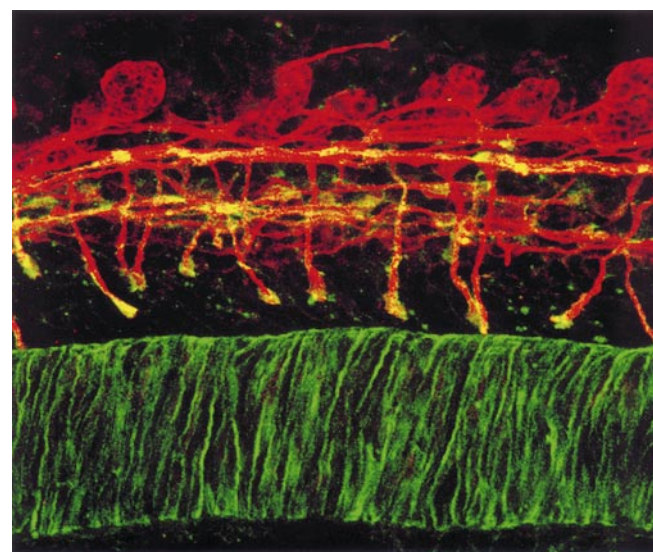
2. Ellis, T. *et al.* *Diabetes* 48 (suppl. 1) Abstr. (1999).

In vivo regulation of axon extension and pathfinding by growth-cone calcium transients

Timothy M. Gomez & Nicholas C. Spitzer

Nature 397, 350–355 (1999)

The cover image on this issue should have been orientated as shown here to conform with the location of the notochord as described in the cover caption. □



The role of mat-forming diatoms in the formation of Mediterranean sapropels

Alan E. S. Kemp, Richard B. Pearce, Itaru Koizumi, Jennifer Pike & S. Jae Rance

Nature 398, 57–61 (1999)

In this Letter, the legend to Fig. 2 has the descriptions switched for the last two panels. The image shown in Fig. 2c is in fact of *H. hauckii*, with scale bar 20 μ m, and that in Fig. 2d is of *P. calcar-avis*, with scale bar 10 μ m.

Genomics and bioinformatics

This issue focuses on equipment and software — including freeware, so read on — for the rapidly expanding genomics and bioinformatics markets.

Internet Bioinformatics Service

From GATC

www.gatc.de

An online bioinformatics service

ConSequence, from GATC, of Konstanz, Germany, can scan your DNA sequences to extract open reading frames (ORFs) which are then the basis for a detailed protein sequence analysis. It provides automated annotation of DNA sequences based on the proven Pedant Pro system (MIPS, Munich). It searches the databases at EMBL, TREMBL, PIR, PDB, Blocks, PROSITE, Pfam, Protfam and GenBank for DNA and protein homologies and displays the results online as browsable text and graphics. An e-mail alert system performs an automatic BLAST search every week for a year on request.

Reader Service No. 100

NetPrimer

From Premier Biosoft www.premierbiosoft.com

A primer analysis program available free

NetPrimer is a web-based program that analyses individual primers or pairs of primers. It is available free from the Premier Biosoft web site (for Windows and Mac). Primer combines primer design algorithms with a web-based interface allowing the user to analyse primers over the Internet. All primers are analysed for melting temperature using the nearest-neighbour thermodynamic for accurate T_m prediction. Primers are analysed for all secondary structures including hairpins, self-dimers, and cross-dimers in primer pairs. Download from www.PremierBiosoft.com/NetPrimer.html.

Reader Service No. 101

Web DoPrimer

From Interactiva Biotechnologie GmbH

www.interactiva.de

Online primer design software on the web

DoPrimer programs are platform independent, based on the Java programming language. A web-enabled version of DoPrimer can be used free over the Internet via the company's site to design single sets of PCR and DNA sequencing primers. Web DoPrimer also offers a search interface to the EMBL sequence database for the retrieval of target sequence information. Primers designed with Web DoPrimer can be ordered instantly over the Internet from Interactiva. The DoPrimer technology is targeted at those researchers who need to design large sets of primers for

File name	Sample description	Sequences	Length	DNA source	Genetic code	Viewable until	Current state
seq3	genomische DNA	1	1756	Eukaryotic genomic DNA (6 frame translation)	Standard	0000-00-00 00:00:00	done
seq2	human cDNA	1	487	Eukaryotic cDNA/EST	Standard	0000-00-00 00:00:00	done
seq1	E. coli DNA	1	4363	Bacterial DNA	Standard	0000-00-00 00:00:00	done

Activate or modify alert requests	Alert expires on:	Alert result:
genomische DNA 0-4	2000-2-24	RESULT

Konstanz companion: this server will keep in touch by e-mail.

PCR, DNA sequencing and Interactiva's proprietary XNA on the Gold chip system. DoPrimer Pro is the next development, due out this summer.

Reader Service No. 102

K6/ODC mouse

From Taconic

www.taconic.com

Transgenic mice keep on coming

Latest transgenic mouse from prolific rodent supplier Taconic is the K6/ODC, used to determine carcinogenicity and to evaluate anticancer drugs. This mouse, produced under licence from ODC Mouse Group, Inc. of Drexel Hill, Pennsylvania, overexpresses ODC (ornithine decarboxylase) in the skin, making it susceptible to tumour development.

Reader Service No. 103

BOSS

From Beckman Coulter

www.beckmancoulter.com

BOSS (Beckman OligoSequence Server) freeware for those producing oligos for other labs

This piece of freeware for the Oligo 1000M



Awaiting instructions from the BOSS.

DNA synthesizer downloads text sequence information into the instrument for order entry and billing documentation. It also serves as a reference database and synthesis log book, and interacts with the Oligo 5 primer design program.

Reader Service No. 104

Hydra Microdispensers

From Robbins Scientific

www.robsci.com

Updated brochure covers microlitre dispensing without the use of pipette tips

Hydra microdispensers are used in genome mapping and sequencing laboratories for dispensing template DNAs and reagents for PCR and cycle sequencing reactions. Pharmaceutical companies use Hydra dispensers for assays used for high-throughput screening of drug libraries. The brochure details the specifications for all Hydra models with recommended applications. Highlighted in this brochure is a systematic study performed by dispensing aqueous and DMSO samples into wet (containing buffer) or dry plates as two ways of measuring dispensing precision.

Reader Service No. 105

Uni-filter 384

From Whatman

www.whatman.com

Filter bottom microplates for high-throughput screening

The robotics-compatible Uni-filter 384-well filter bottom microplates have the same size footprint as the widely available standard 96-well microplate. This means that each 384-well microplate fits the majority of microplate readers that read 384 wells. In

addition to fitting commonly used microplate readers, each filter bottom microplate can be used with any automated liquid-handling and reagent-dispensing systems built to handle 384-well microplates. Each well has a capacity of 100 µl. Drip directors are integral parts of the device and facilitate collection of filtrate without crosstalk.

Reader Service No. 106

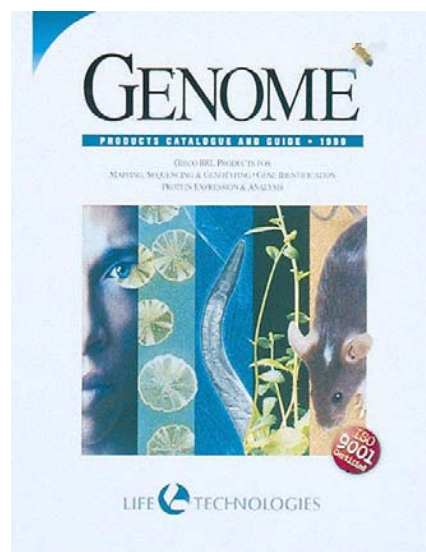
GeneStorm

From Invitrogen www.invitrogen.com

Over 500 expression-ready human gene clones

Invitrogen's growing collection of full-length human genes is optimized for functional genetic and other gene-analysis studies. The clones are designed to save research time, to speed drug discovery and diagnostic development. The GeneStorm product line makes the study of gene function as simple as ordering a gene and performing the desired genetic or biochemical tests as convenient. For researchers, this removes the need to identify, isolate, clone and test a gene for expression.

Reader Service No. 107



For GIBCO BRL, read Life Technologies.

Genome products catalogue

From Life Technologies www.lifetech.com

Detailed information on GIBCO BRL genome products and services

Colour-coded by product technique or application, this new catalogue also features a section on customizable products and services. A reference section includes enzyme applications charts and chapters on molecular biology troubleshooting, *E. coli* transformation and transfection, and how to access the Tech-OnLine and 24-Hour Crisis Response System.

Reader Service No. 108

ChemiGenius

From Syngene www.syngene.co.uk

For fast imaging of chemiluminescence applications

The 16-bit Peltier-cooled CCD camera reduces dark current to almost undetectable levels, say Syngene. ChemiGenius is also available with a high-performance 1,300 × 1,030 pixel resolution camera. The ergonomic light-tight darkroom housing



Tower records: a ChemiGenius at work.

the camera and motorized zoom lens is fully computer-controlled. GeneTools analysis software, included with the system, rapidly performs such tasks as automatic 1D analysis, densitometry, band matching, PCR quantification, colony counting and spot/blots.

Reader Service No. 109

Premium RNA

From Clontech www.clontech.co.uk

A range of panels for gene expression assay

Clontech's new range of tools for gene cloning and the analysis of gene expression are made using the company's Premium RNA — intact and specially pure total and poly(A)⁺ RNA samples isolated from hard-to-obtain tissues. Total RNA Panels, each of which includes Premium Total RNA from six different human tissues, are recommended for RNase protection assays. The Human Digestive System 12-Lane Multiple Tissue Northern (MTN) Blot features Premium Poly A⁺ RNA samples from 12 human digestive-system tissues. These same 12 tissues are represented in the Human Digestive System Multiple Tissue cDNA (MTC) Panel, a set of normalized first-strand cDNAs suitable for RT-PCR analysis of gene expression.

Reader Service No. 110

RNA loading mix

From GenHunter

www.nashville.net/genhut/genhunt.html

A mixture to denature RNA before gel analysis

RNA Loading Mix, available from GenHunter Corporation, is a pH-buffered solution containing the denaturant, tracking dye and ethidium bromide for a convenient one-step preparation of an RNA sample before loading for gel or northern blot analysis. This loading mix solves the problem of apparent RNA degradation. The mixture is claimed to improve northern blot results and take away the additional chore of adding ethidium bromide to your gel or buffer.

Reader Service No. 111

DNA-Pure

From CPG www.cpg-biotech.com

Genomic DNA purification in under an hour

The DNA-Pure Genomic MAXI Kit provides a rapid protocol for the isolation of high-quality, ultra-pure genomic DNA from whole blood, lymphocytes, cultured cells or tissue. Typical yields are approximately 20–40 µg per ml whole blood, and 10–25 µg per 10⁶ cells. Use of the kit avoids the need for organic extractions. Genomic DNA is recovered in yields suitable for PCR, restriction enzyme digestion and

ADVERTISEMENTS

Pepsyn Ltd
High Quality Peptide Synthesis
 'Standard' and 'non-standard'
 Fast delivery
 Competitive prices
 Also advisory services
 Contact: Dr J A Smith, Chairman, Pepsyn Ltd
 School of Biological Sciences, Life Science Building
 The University of Liverpool, Liverpool L69 7ZB, UK
 Tel: 0151 794 4329 Fax: 0151 794 4349
 e-mail: J.A.Smith@liverpool.ac.uk
 Website: www.pepsyn.co.uk

READER ENQUIRY NO. 9

Homogenizers?
 ■ Tissue-Tearor™
 ■ BeadBeater™
 ■ Mini-BeadBeater™
 ■ BioHomogenizer™
Check Us Out
www.biospec.com
 BioSpec Products, Inc.
 Ph/Fax 918-336-3363
 e-mail info@biospec.com
 New Probe Sizes
 0.45cm!

READER ENQUIRY NO. 54

Southern blots. The system is based on a special lysis buffer technology and does not use DNA-binding resins, minimizing DNA shearing. The kit also includes a special lysis solution for use when isolating DNA from whole blood.

Reader Service No. 112

NA 2000

From AutoGen www.autogen.com
A fully automated DNA-RNA isolation system

The Model NA 2000 from AutoGen automates the extraction and purification of genomic DNA or RNA from animal cultured cells and tissues. The system, which is totally automated and requires no manual intervention, produces nucleic acids in yields comparable to those achieved using manual methods. The NA 2000 will isolate DNA from: plant material; double-stranded DNA from BAC, PAC, plasmid, cosmid; single-stranded DNA from M13, λ and yeast artificial chromosome (YAC).

Reader Service No. 113

MasterPure

From Epicentre Technologies
www.epicentre.com

Yeast DNA purification in 30 minutes

The MasterPure yeast DNA purification kit makes use of a rapid protocol eliminating the need for spin columns, hazardous solvents, or resin-based extractions. Simply spin down the yeast liquid culture, add reagent, heat and precipitate. The manufacturers claim that with MasterPure, DNA recoveries up to 15 times those obtained with other DNA extraction methods are realized from 1.5 ml overnight liquid cultures. The extracted DNA is suitable for PCR without further purification. MasterPure is scalable for larger extraction procedures. The kit is available in a 10-reaction trial size and a 200-reaction standard size.

Reader Service No. 114

HS114/114F

From Molecular Research Center
www.mrcgene.com

Hybridization, with or without formamide

MRC's high-efficiency hybridization systems come in two versions, HS 114 and (with formamide) HS 114F. Each is a ready-to-use hybridization solution for northern, Southern and dot/blot hybridization. These systems are designed for maximum enhancement of hybridization signals and low background, making them especially useful for detecting low-abundance mRNAs. MRC claim that the high sensitivity of these systems eliminates the need for

the use of poly(A)⁺ RNA in northern blotting. Only 3–10 μ g of total RNA per lane is required to detect most mRNAs.

Reader Service No. 115

ULTRAhyb

From Ambion www.ambion.com
Hybridization, formamide-based

ULTRAhyb is a formamide-based hybridization solution for northern, Southern and microarrays. Ambion claim a 20–100-fold increase in the sensitivity of DNA probes compared with established methods, as well as decreased exposure time. Targets that once required a 5-day exposure can be achieved in 8 hours. ULTRAhyb is compatible with any format presently in use, including radiolabelled or non-isotopically labelled RNA or DNA probes.

Reader Service No. 116

In situ hybridization

From Hybaid www.hybaidd.co.uk
A new custom in situ hybridization service

This new service is being run by Hybaid, in conjunction with Orion Molecular Services. Tissue preparation, sectioning and mounting can be carried out to order, as well as cDNA cloning and culturing suf-

ficient plasmid for ³⁵S probe preparation.
Reader Service No. 117

GenoSensor

From Vysis www.vysis.com
Commercial genomic array system for detecting abnormal gene numbers

GenoSensor consists of DNA probes on a miniaturized, ready-to-use genomic array, instrumentation and reagents. It uses comparative genomic hybridization (CGH) technology to survey simultaneously multiple gene targets for changes in copy number in a single test. Through comparative DNA binding, or hybridization, genomic changes within the sample DNA are quantitated by analysing the colour ratios of the individual gene targets on the array.

Reader Service No. 118

BioChip Arrayer

From Packard www.packardinst.com
Upgrades to this aid to automation

The BioChip Arrayer is a high-precision, low-volume, automated pipetting platform for research and development of microarray assays. Changes to the pressure management system now allow for greater control during dispensing. Pressure to the PiezoTips is maintained during dispensing by regularly

ADVERTISEMENTS

NEW ANTIBODIES FOR UBIQUITIN AND PROTEASOME RESEARCH FROM AFFINITI

20S Proteasome α - & β -subunits: HC2, HC3, HC8, HC9, ι , XAPC7, ζ , HC10, HN3, Z, MECL1, LMP2, LMP7; 19S regulator ATPase & non-ATPase subunits; 11S regulator subunits: PA28 α , β and γ ; **Signalosome** subunits: PGP9.5; E2; Ub-conjugates

For data sheets or more information on AFFINITI's extensive range of products for ubiquitin and proteasome research visit our website at:

[HTTP://WWW.PROTEASOME.COM](http://www.proteasome.com)

AFFINITI Research Products Ltd., Mamhead Castle, EXETER, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affiniti-res.com>

READER ENQUIRY NO. 38

The Mouse Kit™ for Gene Targeting in ES Cells

- ✓ Complete with embryonic stem (ES) cells, feeder cells and essential reagents (including FBS) for engineering one knockout mouse strain
- ✓ Low-passage AB2.2-Prime ES Cells® (129/SvEv)
- ✓ Uniform, high-quality ESQ feeder cells (mitotically inactivated STO cells) engineered for neo and puro resistance
- ✓ Prequalification of ESQ feeder cells and reagents for use with AB2.2-Prime ES Cells

*Manufactured by Lexicon Genetics Incorporated

www.stratagene.com
Tel: (619) 535-5400 or (800) 424-5444

READER ENQUIRY NO. 6

Detect residual DNA with Wako's DNA Extractor Kit or Isolate Genomic DNA with DNA Extractor WB Kit Both use the Nal Method No Phenol No Chloroform

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 39



Assay, assay: Packard's approach to microassays.

spaced bi-directional syringe pump adjustments and feedback through pressure sensors that allow the instrument to verify dispense accuracy continuously and handle samples of varying viscosity, such as DMSO-solubilized compounds or DNA samples.
Reader Service No. 119

ScanArray 4000/5000

From General Scanning Inc.

www.scanarray.com

Four-colour and custom microarray analysis

General Scanning's latest developments for fluorescent microarray and biochip analysis are the ScanArray 4000 and ScanArray



General improvements: the latest in ScanArrays.

5000. These perform four-colour fluorescent imaging with automated calibration, tracking and analysis of custom or commercially available microarrays. A range of fluorescent dye options can be specified. The systems run in Windows NT. A "build-to-order" option is offered.

Reader Service No. 120

GeneScape Portal

From CuraGen Corporation

www.curagen.com

An Internet Portal and new genomics software

Announced on 3 May, GeneScape Portal provides web-based access to the tools and databases necessary for datamining, managing workflow, and conducting genomics experiments over the Internet. CuraGen is offering secure access to potential partners for the purpose of evaluating these research tools, interfaces, and databases. Through the CuraTools suite of genomics data analysis software, the company is linking information from its proprietary databases with those publicly available, such as the Human Genome Project and the SNP Consortium database. CuraTools is accessible through the GeneScape Portal on www.curagen.com.
Reader Service No. 121

DNA Massarray

From Sequenom

www.sequenom.com

Automating PCR amplification and the primer extension reactions

This DNA analytical system can be operated in conditions typical of large-scale industrial processes. The DNA MassArray APL fully automates PCR amplification and primer extension reactions using laboratory automation co-developed with Robocon. The APL then transfers nanolitre quantities of each sample to SpectroChip for final analysis by a BrukerSequenom MALDI-TOF mass spectrometer.

Reader Service No. 122

ChipReader

From Virtek

www.virtek.ca

DNA microarray imaging system for high throughput

The ChipReader laser confocal system for imaging DNA microarrays offers simulta-

neous reading of up to five different fluorescent dyes. This new imaging system can be used on any microarray format — including standard microscope slides and 3 × 4-inch (76 × 98-mm) glass plates. Software is icon-driven Windows NT.
Reader Service No. 123

GeneTAC

From Genomic Solutions

www.genomicsolutions.com

Image and analyse labelled biochips

The GeneTAC isotopic imager uses direct beta detection imaging and proprietary analysis technologies for rapid and direct quantification of biochips labelled with the commonly used radioactive isotopes, including tritium. Spatial resolution of between 15 and 28 µm is achievable. The GeneTAC Isotopic Imager offers real-time imaging results and optional double labelling detection (for example tritium/phosphorus-32). The system is compatible with most isotopically-labelled biochips up to 32 mm × 24 mm in area.

Reader Service No. 124

CE capillaries

From J&W Scientific

www.jandw.com

Capillary electrophoresis capillaries for ABI PRISM 310 genetic analysers

These replacement CE capillaries for DNA sequencing are preconfigured, bare-fused silica capillaries designed with the latest features including a detection window that is optically clean, scratch- and debris-free. The capillary ends are precision cut to minimize the risk of sample injection errors. Available in economy packs of 4 and 8 capillaries, CE capillaries come in two lengths (0.05 mm ID × 47 cm and 0.05 mm ID × 61 cm).

Reader Service No. 125

Gene-Machine

From Rosys Anthos

www.rosys-anthos.com

'Walk-away' preparation of purified genomic DNA from biological samples

The new Gene-Machine from Rosys Anthos fully automates high-throughput DNA preparations by combining the Plato robotic microplate processor with the FTA Gene Guard System. As well as genomic DNA purification, the system will automate other protocols including PCR, and post-PCR analysis by ELISA. The system will also prepare archive cards to be used for long-term storage of DNA samples.

Reader Service No. 126

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

"Beside good science, there is much humour and much of interest outside the field of science. ... In my case, it kept me awake and delighted until the small hours"

Keith J. Laidler, Professor Emeritus of Chemistry, Ottawa University

The Bedside Book Genius and Eccentricity in Science 1869–1953

Edited by Walter Gratzer, with a foreword by Stephen Jay Gould

A fascinating selection of articles from Nature, offering a panorama of science seen against the background of 19th and 20th century history.



Price: £19.95/US\$29.95 + postage
Discount rate available for current subscribers.
To order, or for more information, contact: Nature Marketing
Tel. +44 (0) 171 843 4985
Fax. +44 (0) 171 843 4998
E-mail: subscriptions@nature.com